Package 'BIPnet'

October 2, 2025

Type Package

Title BIPnet: A Bayesian integrative approach with incorporation of network information

Version 1.0

Date 2024-07-20

Author Thierry Chekouo and Sandra Safo

Maintainer Thierry Chekouo <tchekouo@umn.edu>

Description The BIPnet package implements an MCMC algorithm for a Bayesian analysis that integrates multi-omics data types, clinical covariates and clinical response variables.

SystemRequirements GSL (GNU Scientific Library)

License GPL (>= 2.0)

Imports AUC, MASS, Matrix

RoxygenNote 6.1.0

Encoding UTF-8

R topics documented:

`BIP` *An MCMC algorithm to perform a Bayesian integrative approach for multi-omics data.*

Description

The method can incorporate clinical covariates, clinical response variables and external grouping network information. It performs two methods: i) BIPnet: A Bayesian approach that measures the association between multiple data types and incorporates external prior grouping information; ii) BIP does not account for external grouping information. The algorithm computes (i) the marginal posterior probabilities (MPP) of inclusions for each component and each feature within active components; (ii) posterior mean of the share latent components; iii) Estimated Loading obtained using a threshold on the MPP of inclusions for each feature; iv) Posterior means of group effects.

Usage

```
BIP(dataList=dataList,IndicVar=IndicVar, groupList=NULL,Method=Method,nbrcomp=4,
   sample=5000, burnin=1000,nbrmaxmodels=50,priorcompselv=c(1,1),priorcompselo=c(1,1),
   priorb0=c(2,2),priorb=c(1,1),priorgrpsel=c(1,1),probvarsel=rep(0.05,length(dataList)),
chainNbr=1)
```

Arguments

dataList	List of data sets: If M is the number of omics data types, then this list is of length M+1. Otherwise, if we include the covariate set, then its of length M+2. Each omics dataset object is a data matrix measured in a set of n samples. For each omics matrix, the rows represent samples, and the columns represent genomic features.
IndicVar	This outputs indicates which type of variables. For inctance, if IndicVar=c(0,0,1,2), the first two variable are omics data that would require variable selection, the third variable is the response variable and the fourth is the set of clinical covariates.
groupList	List of groups of length M for each omics data. For each omics data type l, group is a matrix $P_l \times K_l$ where P_l is the number of features, K_l is the number of groups and each entry is either 1 or 0, with 1 if the feature belongs to the corresponding group. If Method="BIP", then groupList=NULL, the default value.
Method	Method is either BIPnet (with network information) or BIP (withouth network information)
nbrcomp	Number of components of the shared components
sample	Total number of MCMC draws. It must be larger than burnin.
burnin	Number of draws to discard for burn-in
nbrmaxmodels	Number maximum of models for prediction
priorcompselv	Hyperparameters of a beta(a,b) distribution, prior distribution of the probability of selecting components
priorcompselo	Hyperparameters of a beta(a,b) distribution, prior distribution of the probability of selecting a component associated with the response.
priorb0	Hyperparameters of a gamma distribution; prior distribution of the paramater b_{l0} that controls the shrinkage of loadings when there is no grouping information
priorb	Hyperparameters of a gamma distribution; prior distribution of group effect coefficients b_{lj}, $j = 1, ..., K_l$.
priorgrpsel	Hyperparameters of a beta(a,b) distribution, prior distribution of the probability of selecting groups.
probvarsel	Prior probability for variable selection for each data type. It's a vector of size the number of data types. Please note that when a data type m is not subject of variable selection (e.g., covariate variables or outcome), then probvarsel[m]=1.
chainNbr	MCMC chain number. The default is 1 for one MCMC number 1. If you want to run N multiple MCMC chains, make a loop in R as for (i in 1:N) BIP(...,chainNbr=i)

Details

The function will return several R objects, which can be assigned to a variable. To see the results, use the "$" operator.

Value

EstU	Posterior mean of the shared components U. Important components is determined using MPPs of γ
VarSelMean	MPP of feature selection for each omics data type for every component; this computes the proportion of iterations that a variable or feature is selected in each component $l = 1, ..., r$
VarSelMeanGlobal	MPP of feature selection for each omics data type; this computes the proportion of iterations that a variable or feature is selected in at least one component.
CompoSelMean	MPPs of the selection of components l.
GrpSelMean	MPPs of the selection of groups if BIPnet is chosen.
GrpEffectMean	Posterior mean of the group effects
....	The other parameters are used for prediction

References

Thierry Chekouo and Sandra Safo (2023), Bayesian Integrative Analysis and Prediction with Application to Atherosclerosis Cardiovascular Disease, *Biostatistics*, , Volume 24, Issue 1.

See Also

BIPpredict

Examples

```
### We run Scenario 1-Setting 1 in the manuscript
library(BIPnet);
dat=Simulate(scenario=1,setting=1,overlap="yes",seed=1);

dataList=list(dat$X1,dat$X2,dat$Y)
PathList=list(dat$Group1,dat$Group2)
BA=BIP(dataList=dataList,IndicVar=c(0,0,1),groupList=PathList,Method="BIPnet",nbrcomp=4,
       sample=5000, burnin=1000, probvarsel=rep(0.05,length(dataList)));
TrueVar1=dat$TrueVar1;
IndicVar=BA$IndicVar;
Np=length(IndicVar)
for (m in 1:Np){
if (IndicVar[m]==0){
print(ComputeVarCriteria(BA$VarSelMeanGlobal[[m]],TrueVar1))
}
}
```

BIPpredict *Prediction of the response variable and shared components*

Description

This function computes predicted response and shared components on a test data set.

Usage

```
BIPpredict(dataListNew=dataListNew,Result=Result,meth="BMA")
```

Arguments

dataListNew	List of new data sets: Do not include the response as you may not have be available for a new set of samples.
Result	This is a BIP object obtained using function BIP on the training set.
meth	meth is either "BMA" (Bayesian model averaging) or "noBMA" for no BMA. In this case, the prediction is performed using only one model obtained using threshold of 0.5 on the MPPs

Details

The function will return both predicted responses and predicted shared latent components on a test data.

References

Thierry Chekouo and Sandra Safo (2023), Bayesian Integrative Analysis and Prediction with Application to Atherosclerosis Cardiovascular Disease, *Biostatistics*, , Volume 24, Issue 1.

Examples

```
### We run Scenario 1-Setting 1 in the manuscript
library(BIPnet);
dat=Simulate(scenario=1,setting=1,overlap="yes",seed=1);
dataList=list(dat$X1,dat$X2,dat$Y)
PathList=list(dat$Group1,dat$Group2)
BA=BIP(dataList=dataList,IndicVar=c(0,0,1),groupList=PathList,Method="BIPnet",nbrcomp=4,
       sample=5000, burnin=1000);
#### Prediction performance
datTest=Simulate(scenario=1,setting=1,overlap="yes",seed=2);
X_new=list(datTest$X1,datTest$X2);
Pred=BIPpredict(dataListNew=X_new,Result=BA,meth="BMA")
### Compute mse

y_new=datTest$Y;
mse=mean((Pred$ypredict-as.vector(y_new))^2);
mse
```

ComputeVarCriteria *Compute criteria to evaluate variable selection performamce*

Description

This function computes false negative rates (FNR), false positive rate (FPR), F1-measure (F1) and AUC.

Usage

```
ComputeVarCriteria(pred.prob,truegroups,thres=0.5)
```

Arguments

pred.prob	(Predicted) probabilities (e.g. marginal posterior probabilities)
truegroups	Binary vector with 1 as positive and 0 as negative.
thres	Threshold on the probablities used to compute FPR, FNR and F1. Default is 0.5.

Details

The function returns FPR, FNR, F1 and AUC.

References

Thierry Chekouo and Sandra Safo (2023), Bayesian Integrative Analysis and Prediction with Application to Atherosclerosis Cardiovascular Disease, *Biostatistics*, , Volume 24, Issue 1.

Simulate *Generation of simulated data as explained in the reference manuscript.*

Description

This function generates data described in the manuscript.

Usage

```
Simulate(scenario=1,setting=NULL,overlap=NULL,seed=1)
```

Arguments

scenario	We have three scenarios as decrisbed in the paper. Scenario 1 has 5 settings, and scenarios 2 and 3 have both two choices: overlap and no overlap.
setting	Settings for secnario 1 only. It's a total of 4 settings
overlap	If yes, then there are some overlap between components, otherwise, there is no overlap. This input only works when scenario is either 2 or 3.
seed	Seed to generate random numbers.

Details

The function will return two data matrix X1 and X2, an estimated response Y, shared component matrix U, loadings A1 and A2, and the grouping information (Group1 and Group2)

References

Thierry Chekouo and Sandra Safo (2023), Bayesian Integrative Analysis and Prediction with Application to Atherosclerosis Cardiovascular Disease, *Biostatistics*, , Volume 24, Issue 1.

Examples

```
dat=Simulate(scenario=3,setting=NULL,overlap="no",seed=1)
str(dat)
```

Index