

Package ‘HMRFBayesHiC’

February 3, 2015

Type Package

Title HMRFBayesHiC conduct Hidden Markov Random Field (HMRF) Bayes Peak Calling Method on HiC Data

Version 1.0

Date 2015-01-30

Author Zheng Xu

Maintainer Zheng Xu <xuzheng@email.unc.edu>

Description we propose a hidden-Markov random field based Bayesian peak caller to explicitly model the spatial dependency of chromatin interaction peaks. This package is to implement our HMRF-Bayes method.

License GPL-3

Depends colorRamps, fields

R topics documented:

HMRFBayesHiC-package	2
HiC.Toy.Data	3
HiC_HMRF_Bayes	4
HiC_HMRF_Bayes_File	5
HiC_Output	8
HiC_Plot_Heatmap	9
HiC_Prepare_Matrix	10
Index	12

HMRFBayesHiC-package

Hidden Markov Random Field Bayesian Hi-C Peak Calling

Description

HMRFBayesHiC helps researchers to conduct peak calling of Hi-C data by our Hidden Markov Random Field-Bayesian method. The input data is a text file, with 4 columns respectively as fragment 1, fragment 2, observed frequency and expected frequency. The output peak probability result is a text file, with 5 columns respectively as fragment 1, fragment 2, observed frequency, expected frequency and peak probability. HMRFBayesHiC include two ways for the user to conduct peak calling. One way is the pipeline calling via the function `HiC_HMRF_Bayes`. The user just need to specify input data file and output result file and `HiC_HMRF_Bayes` will conduct all things. The other way is the three-step running, using `HiC_Prepare_Matrix`, `HiC_HMRF_Bayes` and `HiC_Output` respectively for data loading, peak calling and result outputting steps. The demonstration code is listed below and the toy data include Hi-C information for Fragment 15060-15069 as the file "toyHiC_frag15060_15069.txt".

Details

```
Package:  HMRFBayesHiC
Type:     Package
Version:  1.0
Date:     2015-01-30
License:  GPL-3
```

The pipeline function `HiC_HMRF_Bayes` conducts our HMRFBayes peak calling from input file to output file, in three steps: (1) data loading using `HiC_Prepare_Matrix`, (2) peak calling using `HiC_HMRF_Bayes` and (3) results outputting using `HiC_Output`.

Author(s)

Zheng Xu

Maintainer: Zheng Xu <xuzheng@email.unc.edu>

References

Xu et al, A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interactions in Hi-C Data, Manuscript.

Examples

```
## An example of loading data, run peak caller and output results.
## Our toy example of HiC data is for fragment 15060-15069 of IMR90 12 cell data.

## Conduct Hi-C HMRF in three steps.
## (1) loading data, (2) peak calling and (3) outputting results.
HiC.Data=HiC_Prepare_Matrix(file="toyHiC_frag15060_15069.txt",
                             verbose=TRUE, sep="\t", header=TRUE)
HiC.Results=HiC_HMRF_Bayes(HiC.Data$ObservedCountMatrix,
```

```

HiC.Data$ExpectedCountMatrix, burnIn = 5000, iter = 5000, verbose=1)
HiC_Output(HiCResults=HiC.Results, fragIDs=HiC.Data$fragIDs,
  probfile="toyHiC_frag15060_15069_Results.txt",
  estimatorfile="toyHiC_frag15060_15069_Estimated_Parameters.txt")

## Conduct Hi-C HMRF in one function for all.
HiC_HMRF_Bayes_File(file="toyHiC_frag15060_15069.txt",
  probfile="toyHiC_frag15060_15069_Results.txt",
  estimatorfile="toyHiC_frag15060_15069_Estimated_Parameters.txt",
  sep="\t",header=TRUE, verbose=TRUE, burnIn = 5000, iter = 5000)

## Plot heatmap of observed frequency, expected frequency and peak probability

## Plot the heatmap of observed frequency
## which is at column 3 of the HiC data file.
HiC_Plot_Heatmap(file="toyHiC_frag15060_15069.txt",index=c(1,2,3),
  verbose=TRUE,sep="\t",header=TRUE,negative.log10.scale=FALSE,
  xlab="Fragment Index",ylab="Fragment Index",main="Observed Frequency")

## Plot the heatmap of expected frequency
## which is at column 4 of the HiC data file.
HiC_Plot_Heatmap(file="toyHiC_frag15060_15069.txt",index=c(1,2,4),
  verbose=TRUE,sep="\t",header=TRUE,negative.log10.scale=FALSE,
  xlab="Fragment Index",ylab="Fragment Index",main="Expected Frequency")

## Plot the heatmap of posterior probability
## which is at column 5 of the HiC result file.
HiC_Plot_Heatmap(file="toyHiC_frag15060_15069_Results.txt",index=c(1,2,5),
  verbose=TRUE,sep="\t",header=TRUE,negative.log10.scale=FALSE,
  xlab="Fragment Index",ylab="Fragment Index",main="Posterior Probability")

```

HiC.Toy.Data

HiC Toy Data: IMR90 Combined 12 Cells, Fragment 15060-15069.

Description

This is the toy data. It include the Hi-C data for Fragment 15060-15069

Usage

```
data(HiC.Toy.Data)
```

Format

A data frame with 45 observations on the following 4 variables.

frag1 a numeric vector. The first fragment ID

frag2 a numeric vector. The second fragment ID

observeCount Observed frequencies

expectCount Expected frequencies

Details

This is the toy data. It include the Hi-C data for Fragment 15060-15069 for combined 12 IMR90 cells in Jin et al (2014)'s paper.

Source

The data may be downloaded from HiView browser. http://www.unc.edu/~xuzheng/HiCBrowser/Part6_HiC_Browser/

References

Jin et al (2014), Nature, A high-resolution map of the three-dimensional chromatin interactome in human cells.

Examples

```
# to show it.
data(HiC.Toy.Data)
```

HiC_HMRF_Bayes

Hi-C HMRF Bayes Peak Caller Based On R Data Matrix

Description

The function of running HiC HMRF Bayes peak caller based on expected frequency matrix and observed frequency matrix.

Usage

```
HiC_HMRF_Bayes(U, E, burnIn = 5000, iter = 5000, verbose = 1)
```

Arguments

U	observed frequency matrix, which is a square matrix. Only upper-triangle entries are used. If an entry in the upper triangle is NA, the function will fetch the entry from its corresponding lower triangle. If both entries are NA, the function will output the message that "Observed frequencies can not contain NAs" and return NULL to indicate false data format.
E	expected frequency matrix, which is a square matrix. Only upper-triangle entries are used. If an entry in the upper triangle is NA or 0, the function will fetch the entry from its corresponding lower triangle. If both entries are NA or 0, the function will output the message that Expected frequencies can not contain NAs or 0's and return NULL to indicate false data format.
burnIn	the number of simulations not used in final results averaging as the BurnIn in Bayesian method.
iter	the number of simulations after the burn-ins. For example, if burnin=4000 and iter=8000, out HiC_HMRF_Bayes will run 12000 simulations.
verbose	a logical variable controlling whether the intermediate results are reported for each 200 simulations. If true, the function will report the estimated parameters in each 200 simulations. If false, not reported.

Details

The function of running Hi-C HMRF Bayes peak caller based on expected frequency matrix and observed frequency matrix

Value

A list including three items.

`posterior.probability`

The estimated peak probability matrix by our HMRF-Bayes method.

`expected.count`

The matrix of expected frequencies.

`observed.count`

The matrix of observed frequencies.

`theta.estimate`

The value of estimated parameter theta.

`phi.estimate` The value of estimated parameter phi.

`psi.estimate` The value of estimated parameter psi.

Author(s)

Zheng Xu

References

Xu et al, A hidden Markov random field based Bayesian method for the detection of long_range chromosomal interactions in Hi_C Data, Manuscript.

Examples

```
## An example of loading data, run peak caller and output results.
## Our toy example of HiC data for fragment 15060-15069.
HiC.Data=HiC_Prep_Matrix(file="toyHiC_frag15060_15069.txt",
  verbose=TRUE, sep="\t", header=TRUE)
HiC.Results=HiC_HMRF_Bayes(HiC.Data$ObservedCountMatrix,
  HiC.Data$ExpectedCountMatrix,
  burnIn = 5000, iter = 5000, verbose=1)
HiC_Output(HiCResults=HiC.Results, fragIDs=HiC.Data$fragIDs,
  probfile="toyHiC_frag15060_15069_Results.txt",
  estimatorfile="toyHiC_frag15060_15069_Estimated_Parameters.txt")
```

Description

This is the pipeline function helps researchers conduct Hi-C HMRF Bayes peak calling from loading input data file to outputting results in one function. It performs three steps: (1) load Hi-C data by the function `HiC_Prep_Matrix`, (2) conduct Hi-C HMRF Bayes peak calling method by the function `HiC_HMRF_Bayes`, and (3) write Hi-C results of estimated peak probabilities and parameters respectively into different text files by the function `HiC_Output`. The input file is specified by `file` parameter, which is a text file, with 4 columns respectively as fragment 1, fragment 2, observed frequency and expected frequency. The prob result file is specified by `probfile` parameter, which is a text file, with 5 columns respectively as fragment 1, fragment 2, observed frequency, expected frequency and peak probability. The estimator result file is specified by "`estimatorfile`" parameter, which is a text file including values of estimated ϕ , θ and ψ in our model specification.

Usage

```
HiC_HMRF_Bayes_File(file, probfile, estimatorfile,
  sep = "\t", header = TRUE, verbose = TRUE, burnIn = 5000, iter = 5000)
```

Arguments

<code>file</code>	the file to be loaded, which is a text file, with 4 columns respectively as fragment 1, fragment 2, observed frequency and expected frequency.
<code>probfile</code>	the file to output the estimated peak probability matrix by HiC-HMRF-Bayes method, either a character string naming a file or a connection open for writing. "" indicates output to the console. The prob file is a text file with 5 columns respectively as fragment 1, fragment 2, observed frequencies, expected frequencies and peak probabilities.
<code>estimatorfile</code>	the file to output the estimated parameters of θ , ϕ and ψ by HiC-HMRF-Bayes method, either a character string naming a file or a connection open for writing. "" indicates output to the console.
<code>sep</code>	the field separator character. Values on each line of the file are separated by this character for the data loading step by the function <code>HiC_Prep_Matrix</code> .
<code>header</code>	a logical value indicating whether the file contains the names of the variables as its first line for the data loading step by the function <code>HiC_Prep_Matrix</code> . If missing, the value is determined from the file format: header is set to TRUE if and only if the first row contains one fewer field than the number of columns.
<code>verbose</code>	a logical value controlling whether some information should be outputted. If it is true, the number of HiC records loaded will be reported in data loading step by <code>HiC_Prep_Matrix</code> will be reported and the intermediate results of estimated parameters for each 200 simulations in peak calling by the function <code>HiC_HMRF_Bayes</code> will be reported.
<code>burnIn</code>	the number of burn-ins for the function <code>HiC_HMRF_Bayes</code> called internally. For example, if <code>burnin=4000</code> and <code>iter=8000</code> , out HiC-HMRF-Bayes will run 12000 simulations.
<code>iter</code>	the number of simulations after the burn-ins for the function <code>HiC_HMRF_Bayes</code> called internally. For example, if <code>burnin=4000</code> and <code>iter=8000</code> , out HiC-HMRF-Bayes will run 12000 simulations.

Details

This is the pipeline function helps researchers conduct Hi-C HMRF Bayes peak calling from loading input data file to outputting results in one function. It performs three steps: (1) load Hi-C data by the function `HiC_Prepate_Matrix`, (2) conduct Hi-C HMRF Bayes peak calling method by the function `HiC_HMRF_Bayes`, and (3) write Hi-C results of estimated peak probabilities and parameters respectively into different text files by the function `HiC_Output`. The input file is specified by `file` parameter, which is a text file, with 4 columns respectively as fragment 1, fragment 2, observed frequency and expected frequency. The prob result file is specified by `probfile` parameter, which is a text file, with 5 columns respectively as fragment 1, fragment 2, observed frequency, expected frequency and peak probability. The estimator result file is specified by "`estimatorfile`" parameter, which is a text file including values of estimated ϕ , θ and ψ in our model specification.

When the data does not follow correct data format: (1) not same columns for each line in the input file, (2) observed frequencies contain NA or (3) expected frequencies contain NA or 0 values, the function will stop and output the information, asking the researchers to check data format.

Value

No return values. The function will output the results in the specified `probfile` and `estimatorfile`.

Author(s)

Zheng Xu

References

Xu et al, A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interactions in Hi-C Data, Manuscript.

Examples

```
## An example of loading data, run peak caller and output results.
## Our toy example of HiC data for fragment 15060-15069.

## Conduct Hi-C HMRF in three steps.
## (1) loading data, (2) peak calling and (3) outputting results.
HiC.Data=HiC_Prepate_Matrix(file="toyHiC_frag15060_15069.txt",
verbose=TRUE,sep="\t",header=TRUE)
HiC.Results=HiC_HMRF_Bayes(HiC.Data$ObservedCountMatrix,
HiC.Data$ExpectedCountMatrix, burnIn = 5000, iter = 5000, verbose=1)
HiC_Output(HiCResults=HiC.Results,fragIDs=HiC.Data$fragIDs,
probfile="toyHiC_frag15060_15069_Results.txt",
estimatorfile="toyHiC_frag15060_15069_Estimated_Parameters.txt")

## Conduct Hi-C HMRF in one function for all.
HiC_HMRF_Bayes_File(file="toyHiC_frag15060_15069.txt",
probfile="toyHiC_frag15060_15069_Results.txt",
estimatorfile="toyHiC_frag15060_15069_Estimated_Parameters.txt",
sep="\t",header=TRUE, verbose=TRUE, burnIn = 5000, iter = 5000)
```

HiC_Output

HiC HMRF Bayes Method Result Output

Description

The function of outputting HiC HMRF Bayes peak caller results to files. It outputs estimated peak probability file into probfile and outputs estimated parameters of theta, phi and psi into estimator file.

Usage

```
HiC_Output(HiCResults, fragIDs, probfile, estimatorfile)
```

Arguments

HiCResults	the results of HiC-HMRF-Bayes peak caller, which can take the output of the function <code>HiC_HMRF_Bayes</code> .
fragIDs	the names of fragments in an increasing order, which can take the output from the function <code>HiC_Prepare_Matrix</code> .
probfile	the file to output the estimated peak probability matrix by HiC HMRF Bayes method, either a character string naming a file or a connection open for writing.
estimatorfile	the file to output the estimated parameters of theta, phi and psi by HiC HMRF Bayes method, either a character string naming a file or a connection open for writing.

Details

The function of outputting Hi-C HMRF Bayes peak caller results to files. It outputs estimated peak probability file into probfile and outputs estimated parameters of theta, phi and psi into estimator file.

Value

It does not return any value. The function will write estimated peak probability in the prob file and estimated parameters in estimator file. The prob file is a text file with 5 columns respectively as fragment 1, fragment 2, observed frequencies, expected frequencies and peak probabilities.

Author(s)

Zheng Xu

References

Xu et al, A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interactions in HiC Data, Manuscript.

Examples

```
## An example of loading data, run peak caller and output results.
## Our toy example of HiC data for fragment 15060-15069.
#
HiC.Data=HiC_Prepare_Matrix(file="toyHiC_frag15060_15069.txt",
                             verbose=TRUE, sep="\t", header=TRUE)
HiC.Results=HiC_HMRF_Bayes(HiC.Data$ObservedCountMatrix,
                            HiC.Data$ExpectedCountMatrix, burnIn = 5000, iter = 5000, verbose=1)
HiC_Output(HiC.Results=HiC.Results, fragIDs=HiC.Data$fragIDs,
            probfile="toyHiC_frag15060_15069_Results.txt",
            estimatorfile="toyHiC_frag15060_15069_Estimated_Parameters.txt")
```

HiC_Plot_Heatmap	<i>HiC Plot Heatmap From File</i>
------------------	-----------------------------------

Description

HiC_Plot_Heatmap load data into the memory from file, plot the heatmap based on score.

Usage

```
HiC_Plot_Heatmap=function(file, index=c(1,2,3), sep = "\t",
                           header = TRUE, verbose = TRUE, negative.log10.scale=FALSE,
                           xlab = "", ylab = "", main = "", cex.main=1, cex.lab=1,
                           col=matlab.like(100))
```

Arguments

file	the file to be loaded, which is a text file, separated by the separator.
index	the column index to be loaded for fragment 1, fragment 2 and score. For example, index=c(2, 3, 4) means column 2,3,4 are respectively for fragment 1, fragment 2 and score.
sep	the field separator character. Values on each line of the file are separated by this character.
header	a logical value indicating whether the file contains the names of the variables as its first line. If missing, the value is determined from the file format: header is set to TRUE if and only if the first row contains one fewer field than the number of columns.
verbose	a logical value, controlling whether the function be verbose in loading the data, reporting the number of observations loaded or be silent (not report).
negative.log10.scale	a logical value, controlling whether to apply the negative log10 transformation to the data. This usually applies for the heatmap of p value in the test.
x.lab	the label of x axis.
y.lab	the label of y axis.
main	the label of the plot title.
cex.main	the font size of the label of the plot title.
cex.lab	the font size of the label of the x axis and y axis.

`col` the gradient color used in heatmap. The gradient color is from `colorRamps` package, which can take `blue2red(n)`, `blue2green(n)`, `green2red(n)`, `blue2yellow(n)`, `cyan2yellow(n)`, `magenta2green(n)`, `matlab.like(n)`, `matlab.like2(n)`, `blue2green2red(n)`, `primary.colors(n, steps = 3, no.white = TRUE)`, `rgb.tables` and `table.ramp`. For more details, please refer to `colorRamps` package.

Details

`HiC_Plot_Heatmap` load data into the memory from file, plot the heatmap based on score.

Value

No return values. The function will report error message if get some errors.

Author(s)

Zheng Xu

References

Xu et al, A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interactions in Hi-C Data, Manuscript

Examples

```
## For our toy example of HiC data for fragment 15060-15069.

## Plot the heatmap of observed frequency
## which is at column 3 of the HiC data file.
HiC_Plot_Heatmap(file="toyHiC_frag15060_15069.txt", index=c(1,2,3),
  verbose=TRUE, sep="\t", header=TRUE, negative.log10.scale=FALSE,
  xlab="Fragment Index", ylab="Fragment Index", main="Observed Frequency")

## Plot the heatmap of expected frequency
## which is at column 4 of the HiC data file.
HiC_Plot_Heatmap(file="toyHiC_frag15060_15069.txt", index=c(1,2,4),
  verbose=TRUE, sep="\t", header=TRUE, negative.log10.scale=FALSE,
  xlab="Fragment Index", ylab="Fragment Index", main="Expected Frequency")

## Plot the heatmap of posterior probability
## which is at column 5 of the HiC result file.
HiC_Plot_Heatmap(file="toyHiC_frag15060_15069_Results.txt", index=c(1,2,5),
  verbose=TRUE, sep="\t", header=TRUE, negative.log10.scale=FALSE,
  xlab="Fragment Index", ylab="Fragment Index", main="Posterior Probability")
```

HiC_Prep_Matrix *HiC HMRF Bayes Method Data Input*

Description

`HiC_Prep_Matrix` load data into the memory and prepare the matrix of expected count and observed count for further analysis.

Usage

```
HiC_Prepare_Matrix(file, verbose = TRUE, sep = "\t", header = TRUE)
```

Arguments

<code>file</code>	the file to be loaded, which is a text file, with four columns respectively as fragment 1, fragment 2, observed frequency and expected frequency.
<code>verbose</code>	a logical value, controlling whether the function be verbose in loading the data, reporting the number of observations loaded or be silent (not report).
<code>sep</code>	the field separator character. Values on each line of the file are separated by this character.
<code>header</code>	a logical value indicating whether the file contains the names of the variables as its first line. If missing, the value is determined from the file format: header is set to TRUE if and only if the first row contains one fewer field than the number of columns.

Details

`HiC_Prepare_Matrix` load data into the memory and prepare the matrix of expected count and observed count for further analysis. If the file specified does not exist, it will output "File not exist". If success, it will return the loaded data. If fails, it will return NULL. Please use `is.null()` to check whether the function runs successfully.

Value

A list including three items.

`ExpectedCountMatrix`

The matrix of expected frequencies.

`ObservedCountMatrix`

The matrix of observed frequencies.

`fragIDs`

The fragment names in an increasing order.

Author(s)

Zheng Xu

References

Xu et al, A hidden Markov random field based Bayesian method for the detection of long-range chromosomal interactions in Hi-C Data, Manuscript

Examples

```
## For our toy example of HiC data for fragment 15060-15069.
HiC.Data=HiC_Prepare_Matrix(file="toyHiC_frag15060_15069.txt",
  verbose=TRUE, sep="\t", header=TRUE)
```

Index

*Topic **HMRFBayesHiC-package**

HMRFBayesHiC-package, [2](#)

*Topic **HiC_HMRFBayes_File**

HiC_HMRFBayes_File, [5](#)

*Topic **HiC_HMRFBayes**

HiC_HMRFBayes, [4](#)

*Topic **HiC_Output**

HiC_Output, [8](#)

*Topic **HiC_Plot_Heatmap**

HiC_Plot_Heatmap, [9](#)

*Topic **HiC_Prepare_Matrix**

HiC_Prepare_Matrix, [10](#)

*Topic **datasets**

HiC.Toy.Data, [3](#)

*Topic **package**

HMRFBayesHiC-package, [2](#)

HiC.Toy.Data, [3](#)

HiC_HMRFBayes, [4](#)

HiC_HMRFBayes_File, [5](#)

HiC_Output, [8](#)

HiC_Plot_Heatmap, [9](#)

HiC_Prepare_Matrix, [10](#)

HMRFBayesHiC

(*HMRFBayesHiC-package*), [2](#)

HMRFBayesHiC-package, [2](#)