Package 'classSNitch'

October 9, 2015

Title Classifies RNA Structure Change in Chemical Mapping Data

Version 0.12.0

Author Chanin Tolson [aut, cre]

Maintainer Chanin Tolson <ctolson@email.unc.edu>

URL http://www.unc.edu/~ctolson/classSNitch/classSNitch.html

Description Mutations in RNA will create a riboSNitch, if important structural elements are disrupted.Recent ultra-high throughput techniques, such as SHAPE-MaP and PARS, enable the collection of structural RNA information on a genome-wide scale. With the ability to gather genome-wide structural information on RNA, it is important to accurately classify these structural data in order to identify those structural changes that result in a phenotypic outcome. We have developed an automated approach to classify structure change in SHAPE data. This method utilizes random forest classification on a set of pattern and magnitude parameters from the mutate and map SHAPE data set (or another user specified data set) to build a classifier. The classifier is then used to identify structure change in other SHAPE traces. Enabling scientists to identify structure change may help guide experiments that examine RNA structure and its role in biological processes.

Depends R (>= 3.2.2)

License GPL (>= 2)

LazyData true

Imports randomForest,
ROCR,
gplots,
dtw

R topics documented:

`classifyRNA` *classifyRNA*

Description

A function to build a classifier for RNA structure change

Usage

```
classifyRNA(data=NULL, classes=1, cutoff=NULL)
```

Arguments

`data`	Optional data to build the classifier. Default is pre-loaded data.
`classes`	An optional number indicating which class style to use. Only used when data is not supplied. Default is 1.
`cutoff`	An optional vector of length equal to number of classes. The winning class for an observation is the one with the maximum ratio of proportion of votes to cutoff. Default is 1/k where k is the number of classes (i.e., majority vote wins).

Details

This function builds a random forest classifier for RNA structure change in SHAPE data

This function builds a random forest classifier for RNA structure change using the randomForest package.

Value

A classifyRNA object, based on randomForest object (see randomForest package)

call The original call to randomForest

type One of regression, classification, or unsupervised.

predicted The predicted values of the input data based on out-of-bag samples.

importance A matrix with nclass + 2 columns. The first nclass columns are the class-specific measures computed as mean descrease in accuracy. The nclass + 1st column is the mean descrease in accuracy over all classes. The last column is the mean decrease in Gini index.

importanceSD The standard errors of the permutation-based importance measure. A p by nclass + 1 matrix corresponding to the first nclass + 1 columns of the importance matrix.

ntree Number of trees grown.

mtry Number of predictors sampled for spliting at each node.

forest A list that contains the entire forest

err.rate Vector error rates of the prediction on the input data, the i-th element being the (OOB) error rate for all trees up to the i-th.

confusion The confusion matrix of the prediction (based on OOB data).

votes A matrix with one row for each input data point and one column for each class, giving the fraction or number of (OOB) votes from the random forest.

oob.times Number of times cases are out-of-bag (and thus used in computing OOB error estimate)

proximity A matrix of proximity measures among the input (based on the frequency that pairs of data points are in the same terminal nodes).

Note

Organization of the data file: header=TRUE, tab-delimited .txt file

- "column 1" class label
- "column 2" magnitude change
- "column 3" pattern change
- "column 4" change distance
- "column 5" time warping
- "column 6" trace difference
- "column 7" rna length

Options for classes:

- "1" none v. local/global
- "2" global v. none/local
- "3" local v. global

The default data has been gathered from the RNA Mapping Database mutate and map experiments.

Author(s)

Chanin Tolson

References

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22 (randomForest package)

RNA Mapping Database

See Also

`getFeatures`

Examples

```
#build classifier
rf = classifyRNA(classes=1)
#get confusion matrix
rf$confusion
```

`classify_default` *classify_default*

Description

classify_default

References

RNA Mapping Database

`classSNitch` *Package for the autonomous classification of RNA structure change*

Description

Mutations in RNA will create a riboSNitch, if important structural elements are disrupted. Recent ultra-high throughput techniques, such as SHAPE-MaP and PARS, enable the collection of structural RNA information on a genome-wide scale. With the ability to gather genome-wide structural information on RNA, it is important to accurately classify these structural data in order to identify those structural changes that result in a phenotypic outcome. We have developed an automated approach to classify structure change in SHAPE data. This method utilizes random forest classification on a set of pattern and magnitude parameters from the mutate-and-map SHAPE data set (or another user specified data set) to build a classifier. The classifier is then used to identify structure change in other SHAPE traces. Enabling scientists to identify structure change may help guide experiments that examine RNA structure and its role in biological processes.

Author(s)

Chanin Tolson

References

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22 (randomForest package)

RNA Mapping Database

See Also

`getFeatures` `classifyRNA`

Examples

```
#get change features
library("ROCR")
library("gplots")

data("shape_ex")
sample_shape = shape_ex
sample = getFeatures(sample_shape[2:nrow(sample_shape),], base=sample_shape[1,], trim=5)

#predict change
data("mutmap")
sample_class = mutmap[,c(1,4:9)]
cr = classifyRNA(sample_class, classes=1)
cr_pred = predict(cr, sample, type="response")

#plot ROC curve
col = 1
data("mutmap")
data = mutmap
data = data[-which(is.na(data[,col]),arr.ind=TRUE),]
predobj = prediction(cr$votes[,1], data[,col])
perfobj = performance(predobj, tpr, fpr)
aucobj = performance(predobj, auc)
plot(perfobj@x.values[[1]], perfobj@y.values[[1]], lwd=2,
     type="l", xlab="Specificity", ylab="Sensitivity")
points(c(-1,2),c(-1,2), col="red", type="l")
text(0.8, 0.2, paste("AUC: ", format(aucobj@y.values, digits=2), sep=""), cex=1)
```

`getChangeDist` *getChangeDist*

Description

A function to get the average distance of change from the disruption site

Usage

```
getChangeDist(sample, point=rep(0,nrow(sample)), base=sample[1,], margin=1, tol=0.1)
```

Arguments

`sample`	A numeric matrix containing values to be compared (e.g. a set of mutant SHAPE traces).
`point`	An optional numeric vector containing the location of the disruption (e.g. the mutation in an RNA)
`base`	An optional numeric vector containing the values to which the samples are to be compared (e.g. a wildtype SHAPE trace). Default is the first trace in sample.
`margin`	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.
`tol`	An optional number indicating the tolerance for the change. Default is 0.1.

Details

This function determines the average distance of change from the disruption site

This function calculates the average distance of change from the disruption using the change in pattern to determine the location of changes.

Value

A numeric vector of change distances.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)
#reduce noise
samp_nreduce = reduceNoise(samp_norm, trim=1, high=4)
#get change distance
loc = getChangeDist(samp_nreduce)
```

---

getFeatures *getFeatures*

Description

A function to get the features for describing RNA structure change. These features can be used in classification of RNA structure change.

Usage

```
getFeatures(sample, base=NULL, margin=1,  norm=T, noise=T, trim=0, high=NULL,
   tol=0.1, point=rep(0,nrow(sample)), window=ncol(sample), outfile=NULL, append=F)
```

Arguments

sample	A numeric matrix containing values to be compared (e.g. a set of mutant SHAPE traces).
base	An optional numeric vector containing the values to which the samples are to be compared (e.g. a wild type SHAPE trace). Default is the first trace in each file.
margin	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.
norm	An optional boolean to normalize the sample. Default is TRUE.
noise	An optional boolean to reduce noise in the sample. Default is TRUE.

trim	An optional number indicating the number of nucleotides to be trimed from the ends. Default is 0.
high	An optional number indicating the reactivity above which reactivities are considered high. Default is third quartile of the sample in each file.
tol	An optional number indicating the tolerance for the change. Default is 0.1.
point	An optional numerical vector indicating the location of disruption (e.g. mutation point)
window	An optional number indicating the number of columns around the disruption to calculate. Default is the entire trace.
outfile	An optional string indicating the name of the output file. The output file will consist of two columns (magnitude change and pattern change). Default will not output a file.
append	An optional boolean to append the file if an outfile is given. Default is FALSE.

Details

This calculates the change in SHAPE reactivity traces.

This function calculates the magnitude correlation coefficient, pattern correlation coefficient, average change distance, dynamic time warping, average trace difference and rna length.

Value

"outmat" A three column numeric matrix for magnitude, pattern, location and timewarp change.

"outfile" An optional output file for the matrix.

Author(s)

Chanin Tolson

See Also

normalize reduceNoise getMagCC getPatternCC getChangeDist getTimeWarping getTraceDiff

Examples

```
#input files
data("shape_ex")
#get features
params = getFeatures(shape_ex, trim=5, outfile="out.txt")
```

---

getMagCC *getMagCC*

Description

A function to get the correlation coefficient between sample magnitudes

Usage

```
getMagCC(sample, base=sample[1,], margin=1)
```

Arguments

sample	A numeric matrix containing values to be compared (e.g. a set of mutant SHAPE traces).
base	An optional numeric vector containing the values to which the samples are to be compared (e.g. a wildtype SHAPE trace). Default is the first trace in sample.
margin	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.

Details

This function compares the magnitude change between samples.

This function calculates the Pearson correlation coefficient between the base vector and each row (or column) in sample.

Value

A numeric vector of correlation coefficients.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)
#reduce noise
samp_nreduce = reduceNoise(samp_norm, trim=1, high=4)
#get magnitude correlation coefficient
mag = getMagCC(samp_nreduce)
```

---

getPatternCC *getPatternCC*

Description

A function to get the correlation coefficient between sample patterns

Usage

```
getPatternCC(sample, base=sample[1,], margin=1, tol=0.1)
```

Arguments

sample	A numeric matrix containing values to be compared (e.g. a set of mutant SHAPE traces).
base	An optional numeric vector containing the value to which the samples are to be compared (e.g. a wild type SHAPE trace). Default is the first trace in sample.
margin	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.
tol	An optional number indicating the tolerance for the change. Default is 0.1.

Details

This function compares the pattern change between samples.

The pattern for a single SHAPE reacitivity trace is the pattern of increase in reactivity or decrease in reactivity between nucleotides. If the change is less than the tolerance value, it is considered a none change. The pattern change value is the Pearson correlation coefficient between the base vector pattern and the pattern of each row (or column) in sample.

Value

A numeric vector of pattern correlation coefficients.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)
#reduce noise
samp_nreduce = reduceNoise(samp_norm, trim=1, high=4)
#get pattern correlation coefficient
pat = getPatternCC(samp_nreduce)
```

---

getTimeWarping *getTimeWarping*

Description

A function to get the dynamic time warping between sample magnitudes

Usage

```
getTimeWarping(sample, base=sample[1,], margin=1)
```

Arguments

sample	A numeric matrix containing values to be compared (e.g. a set of mutant SHAPE traces).
base	An optional numeric vector containing the values to which the samples are to be compared (e.g. a wildtype SHAPE trace). Default is the first trace in sample.
margin	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.

Details

This function compares the timewarp between samples.

This function calculates the dynamic time warping between the base vector and each row (or column) in sample.

Value

A numeric vector of time warping values.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)
#reduce noise
samp_nreduce = reduceNoise(samp_norm, trim=1, high=4)
#get time warping
tw = getTimeWarping(samp_nreduce)
```

---

getTraceDiff *getTraceDiff*

Description

A function to get the average trace difference between samples

Usage

```
getTraceDiff(sample, base=sample[1,], margin=1,
   point=rep(0,nrow(sample)), window=ncol(sample))
```

Arguments

sample	A numeric matrix containing values to be compared (e.g. a set of mutant SHAPE traces).
base	An optional numeric vector containing the values to which the samples are to be compared (e.g. a wildtype SHAPE trace). Default is the first trace in sample.
margin	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.
point	An optional numeric vector containing the location of the disruption (e.g. the mutation in an RNA)
window	An optional number indicating the number of columns around the disruption to calculate. Default is the entire trace.

Details

This function compares the average trace difference between samples.

This function calculates the average trace difference between the base vector and each row (or column) in sample.

Value

A numeric vector of trace differences.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)
#reduce noise
samp_nreduce = reduceNoise(samp_norm, trim=1, high=4)
#get trace difference
tc = getTraceDiff(samp_nreduce)
```

---

mutmap *Consensus classification, magnitude change, and pattern change in example RNAs*

Description

Example dataset with the calculated magnitude correlation coefficient, pattern correlation coefficient, average change distance, dynamic time warping, and average trace difference for a set of RNAs. Features calculated from SHAPE traces from the RNA Mapping Database. Parameters calculated using getFeatures function in classSNitch package. Consensus classification determined using crowd-sourced manual classification.

Usage

```
mutmap
```

Format

```
data.frame: 2670 obs. of  9 variables:
 $ NvLG.response   : num  0 0 NA 0 0 0 NA NA NA 0 ...
 $ NLvG.response   : int  0 0 NA 0 0 0 NA NA NA 0 ...
 $ LvG.response    : int  NA NA NA NA NA NA NA NA NA NA ...
 $ magnitude change: num  0.988 0.993 0.984 0.978 0.988 ...
 $ pattern change  : num  0.82 0.888 0.866 0.894 0.899 ...
 $ change distance : num  0.559 0.633 0.557 0.441 0.482 ...
 $ time warping    : num  0.142 0.114 0.141 0.154 0.113 ...
 $ trace difference: num  2.52 2.45 1.95 2.97 4.23 ...
 $ length          : num  110 110 110 110 110 110 110 110 110 110 ...
```

References

RNA Mapping Database

Examples

```
data("mutmap")
```

---

normalize *normalize*

Description

A Between-Sample Normalization Function

Usage

```
normalize(sample, base=sample[1,], margin=1, outbase=FALSE)
```

Arguments

sample	A numeric matrix containing values to be normalized (e.g. a set of mutant SHAPE traces).
base	An optional numeric vector containing the values to which the sample is to be normalized (e.g. a wild type SHAPE trace). Default is the first trace in sample.
margin	An optional number indicating if sample is organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.
outbase	An optional boolean indicating if the normalized base should be returned. Default is FALSE.

Details

This function performs between-sample normalization.

This function normalizes the average value of the base vector to 1.5. Each row (or column) in sample is then normalized by minimizing the absolute difference between the base and the sample row (or column).

Value

"samp_norm" A normalized numeric matrix with the same dimensions as sample.

"samp_norm" An optional list with two elements: normalized numeric matrix with the same dimensions as sample and a normalized vector the same length as base.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)

#sample data
sample = matrix(sample(1:100), ncol=10)
base = sample(1:100, size=10)
#normalize
samp_norm = normalize(sample, base)
```

---

predict.classifyRNA *predict.classifyRNA*

Description

A function to classify rna structure change

Usage

```
## S3 method for class classifyRNA
predict(object, sample = NULL, resp="prob", ...)
```

Arguments

object	An object of classifyRNA (see classifyRNA function).
sample	An optional matrix of predictors for magnitude correlation coefficient, pattern correlation coefficient, change distance, time warping and trace difference (e.g. output from getFeatures())
resp	An optional string to determine type of return value. Default is "prob".
...	additional arguments for specific methods

Details

This function predicts RNA structure change in SHAPE data

This function predicts RNA structure change in SHAPE data using a random forest classifier.

Value

A matrix of "response", "vote" or "prob" predictions.

"response" Predicted classes (classes with majority vote)

"vote" Vote count fraction (one column for each class and one row for each input)

"class" Class probabilities (one column for each class and one row for each input)

Author(s)

Chanin Tolson

References

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22 (randomForest package)

See Also

`getFeatures` `classifyRNA`

Examples

```
#input data
data("mutmap")
#build classifier
cr = classifyRNA(classes=2)
#get prediction
cr_pred = predict(cr, mutmap[,4:9])
```

`reduceNoise` *reduceNoise*

Description

A function to reduce noise in SHAPE data

Usage

```
reduceNoise(sample, base=sample[1,], margin=1, trim=0, high=boxplot(sample)$stats[4])
```

Arguments

`sample`	A numeric matrix containing reactivity scores to be compared (e.g. a set of mutant SHAPE traces).
`base`	An optional numeric vector containing the reactivity score to which the samples are to be compared (e.g. a wild type SHAPE trace). Default is the first trace in sample.
`margin`	An optional number indicating if the samples are organized by rows or columns, where 1 indicates rows and 2 indicates columns. Default is 1.
`trim`	An optional number indicating the number of nucleotides to be trimmed from the each end. Default is 0.
`high`	An optional number indicating the value above which reactivities are considered high. Default is third quartile of sample.

Details

This function removes peaks that are high in both the sample and the comparison SHAPE traces.

This function reduces the noise in SHAPE data. For positions where both the base vector and the sample row (or column) is above the high value, the position in the sample row (or column) is set equal to that position in the base vector. The function trims the data by setting the ends of sample equal to the ends of the base vector.

Value

A noise reduced numeric matrix with the same dimensions as sample.

Author(s)

Chanin Tolson

See Also

getFeatures

Examples

```
#sample data
sample = matrix(sample(1:100), ncol=10)
#normalize
samp_norm = normalize(sample)
#reduce noise
samp_nreduce = reduceNoise(samp_norm, trim=1, high=4)
```

---

shape_ex *16S rRNA four-way junction wild-type and mutant SHAPE data*

Description

Example SHAPE traces from 16SFWJ_1M7_0001 Mutate and Map experiment in the RNA Mapping Database.

Usage

```
shape_ex
```

Format

```
 num [1:111, 1:110] 18.89 11.57 3.69 13.7 10.68 ...
 - attr(*, "dimnames")=List of 2
  ..$ : chr [1:111] "16SFWJ_1M7_0001_WT" "16SFWJ_1M7_0001_G126C" "16SFWJ_1M7_0001_G127C" "16SFWJ
  ..$ : chr [1:110] "V3" "V4" "V5" "V6" ...
```

References

RNA Mapping Database

Examples

```
data("shape_ex")
```

Index