Package 'classSNitch'

January 7, 2016

Title Classifies RNA Structure Change in Chemical Mapping Data

Version 1.0.0

Author Chanin Tolson [aut, cre]

Maintainer Laederach Lab <laederachlab@gmail.com>

URL http://classsnitch.r-forge.r-project.org

Description Variations in RNA sequence lead to differences in RNA structure called riboSNitches, if important structural elements are disrupted. Recent ultra-high throughput techniques, such as SHAPE-MaP and PARS, enable the collection of structural information on RNAs at a genome-wide scale. With the ability to gather large amounts of structural information, it is important to accurately identify those structural changes that can potentially result in a phenotypic outcome. We have developed an automated approach to identify and classify structure change in chemical mapping data. This method utilizes random forest classification on a set of seven characterizing features. The default mutate and map SHAPE data sets (or another user specified data set) can be used 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,

RoxygenNote 5.0.0.9000

R topics documented:

`classifyRNA` *classifyRNA*

Description

A function to build a random forest classifier for RNA structure change

Usage

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

Arguments

`data`	Optional data to build the classifier. Default is pre-loaded data.
`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 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" pattern correlation
- "column 3" dynamic time warping
- "column 4" contiguousness
- "column 5" magnitude correlation
- "column 6" change variance
- "column 7" eSDC
- "column 8" change range

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()
#get confusion matrix
rf$confusion
```

`classify_default` *classify_default*

Description

A dataset of class labels and features SHAPE-traces from the mutate and map experiments found in the RMDB

Note

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

- "column 1" class label
- "column 2" pattern correlation
- "column 3" dynamic time warping
- "column 4" contiguousness
- "column 5" magnitude correlation
- "column 6" change variance
- "column 7" eSDC
- "column 8" change range

References

RNA Mapping Database

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

Description

Package for the autonomous classification of RNA structure change.

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 = getFeatures(shape_ex[2:nrow(shape_ex),], base=shape_ex[1,], trim=5)

#predict change
data("mutmap")
cr = classifyRNA(mutmap)
cr_pred = predict(cr, sample, type="response")

#plot ROC curve (no change v. local/global change)
data("mutmap")
predobj = prediction(cr$votes[,1], mutmap[,1]==1)
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)
```

---

getChangeRange *getChangeRange*

Description

A function to get the range of change for a SHAPE trace.

Usage

```
getChangeRange(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 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 calculates the range of change positions for each row (or column) in sample.

Value

A numeric vector of trace change ranges.

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 range
cr = getChangeRange(samp_nreduce)
```

---

getChangeVar *getChangeVar*

Description

A function to get the spread or variance of change for a SHAPE trace.

Usage

```
getChangeVar(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 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 calculates the variance of change positions for each row (or column) in sample.

Value

A numeric vector of trace change variances.

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 variance
cv = getChangeVar(samp_nreduce)
```

---

getContiguous *getContiguous*

Description

A function to get the number of stretches of contiguous change.

Usage

```
getContiguous(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 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 calculates the number of stretches of contiguous change between the base vector and each row (or column) in sample.

Value

A numeric vector of counts

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
contig = getContiguous(samp_norm)
```

---

getESDC *getESDC*

Description

A function to get experimental structure disruption coefficient (eSDC).

Usage

```
getESDC(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 calculates the eSDC between the base vector and each row (or column) in sample.

Value

A numeric vector of eSDC 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 trace difference
eSDC = getESDC(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, 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.
`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 function calculates the pattern correlation coefficient, dynamic time warping, contiguousness of change, magnitude correlation coefficient, change of variance, experimental structural disruption coefficient, and change range. These are features used to describe structure change in SHAPE traces.

Value

"outmat" A seven column numeric matrix for pattern correlation coefficient, dynamic time warping, contiguousness of change, magnitude correlation coefficient, change of variance, experimental structural disruption coefficient, and change range

"outfile" An optional output file for the matrix.

Author(s)

Chanin Tolson

See Also

normalize reduceNoise getPatternCC getTimeWarping getContiguous getMagCC getESDC getChangeRange

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 SHAPE trace 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 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 SHAPE trace change 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

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 SHAPE trace 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 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)
```

---

mutmap — *Consensus classification and features for example RNA SHAPE traces. 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.*

Description

Consensus classification and features for example RNA SHAPE traces. 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

An object of class `matrix` with 167 rows and 8 columns.

References

RNA Mapping Database

Examples

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

normalize *normalize*

Description

A between-sample normalization function for SHAPE traces.

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 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 from an existing classifier.

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 predict method

Details

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()
#get prediction
cr_pred = predict(cr, mutmap[,2:8])
```

---

reduceNoise *reduceNoise*

Description

A function to reduce noise in SHAPE data. This function removes peaks that are high in both the base and the comparison SHAPE traces.

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 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 from mutate and map experiments in the RMDB.*

Description

16S rRNA four-way junction wild-type and mutant SHAPE data from mutate and map experiments in the RMDB.

Usage

```
shape_ex
```

Format

An object of class `matrix` with 111 rows and 110 columns.

References

RNA Mapping Database

Examples

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

Index