

Package ‘IdMappingAnalysis’

April 5, 2014

Imports boot, mclust, RColorBrewer, Biobase

Maintainer Alex Lisovich <alex.lisovich@gmail.com>, Roger Day <day01@pitt.edu>

License GPL-2

Title ID Mapping Analysis

LazyData yes

Type Package

Author Alex Lisovich, Roger Day

Description Identifier mapping performance analysis

Version 1.6.0

biocViews Bioinformatics, Annotation, MultipleComparisons

Date 2013-05-24

Depends R (>= 2.14), R.oo (>= 1.13.0), rChoiceDialogs

Collate 'bootstrap.R' 'corr.R' 'corrData.R' 'data.R' 'dataFilter.R'
'zzz.R' 'display.R' 'idMap.R' 'idMapBase.R' 'idMapCounts.R'
'fit2clusters.R' 'expectedUtility.R' 'idMapDiff.R'
'idMapDiffCounts.R' 'IdMappingAnalysis-package.R' 'jointIdMap.plots.R' 'jointIdMap.R'
'jointUniquePairs.extensions.R' 'jointUniquePairs.interactive.plots.R'
'jointUniquePairs.plots.R' 'jointUniquePairs.R' 'misc.R' 'mixture.R' 'subset.R' 'uniquePairs.R'

R topics documented:

Bootstrap	2
Corr	4
CorrData	5
DataFilter	7
Display	8
examples	9

expectedUtility	10
fit2clusters	11
IdMap	13
IdMapBase	14
IdMapCounts	16
IdMapDiff	17
IdMapDiffCounts	19
JointIdMap	21
JointUniquePairs	22
Misc	24
Mixture	25
Subset	27
UniquePairs	28

Index	31
--------------	-----------

Bootstrap	<i>The Bootstrap class</i>
-----------	----------------------------

Description

Package: IdMappingAnalysis

Class Bootstrap

Object

```

~~|
~~+--IdMapBase
~~~~~|
~~~~~+--Bootstrap

```

Directly known subclasses:

```

public static class Bootstrap
extends IdMapBase

```

The Bootstrap object encapsulates a data frame containing the unique pairs in the first two columns and the correlation results, sd and bias obtained from the bootstrapping procedure in the next 3 columns. During the object creation, the bootstrapping procedure is applied to each row of the experiment set pairs from the CorrData object optionally applying the Fisher transform to the correlation data.

Usage

```
Bootstrap(corrData=NULL, Fisher=FALSE, R=200, verbose=FALSE, ...)
```

Arguments

<code>corrData</code>	CorrData object on which the correlation related bootstrapping is performed.
<code>Fisher</code>	If <code>TRUE</code> , the Fisher transform of data is performed during bootstrapping. Default is <code>FALSE</code> .
<code>R</code>	The number of bootstrap replicates. Default is 200.
<code>verbose</code>	if <code>TRUE</code> enables diagnostic messages. Default is <code>FALSE</code> .
<code>...</code>	Not used.

Value

A Bootstrap object encapsulating the `data.frame` with following columns:

<code>column 1</code>	the first component (primary IDs) of unique pairs. The column name corresponds to the primary key of a source ID Map
<code>column 2</code>	the second component (secondary IDs) of unique pairs. The column name corresponds to the secondary key of a source ID Map
<code>'corr'</code> column	contains the correlation values obtained from bootstrapping
<code>'sd'</code> column	contains the correlation sd values obtained from bootstrapping
<code>'bias'</code> column	contains the correlation bias values obtained from bootstrapping

Fields and Methods**Methods:**

`plot` Scatterplot of bootstrapped results: sd vs correlation .

Methods inherited from IdMapBase:

`[], aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey`

Methods inherited from Object:

`$/, $<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save`

Author(s)

Alex Lisovich, Roger Day

Examples

```
bootstrap<-Bootstrap(examples$corrData,R=20,verbose=TRUE);
class(bootstrap);
bootstrap[1:10,];
```

Corr

*The Corr class***Description**

Package: IdMappingAnalysis

Class Corr**Object**

```

~~|
~~+---IdMapBase
~~~~~|
~~~~~+---Corr

```

Directly known subclasses:

```

public static class Corr
extends IdMapBase

```

Create the Corr object by performing correlations on the CorrData object using the correlation algorithm defined by the method argument. The Corr object encapsulates a `data.frame` containing three columns: the first two are unique pairs and the third is a correlation results with a column name reflecting the correlation method ('pearson', 'spearman' or 'kendall').

Usage

```
Corr(corrData=NULL, method="pearson", verbose=FALSE, ...)
```

Arguments

<code>corrData</code>	CorrData object on which correlation is performed or a <code>data.frame</code> compliant with the Corr object internal data frame format.
<code>method</code>	Correlation method ('pearson', 'spearman' or 'kendall'). Default is 'pearson'.
<code>verbose</code>	if <code>TRUE</code> enables diagnostic messages. Default is <code>FALSE</code> .
<code>...</code>	Not used.

Fields and Methods**Methods:**

<code>getData</code>	Extract correlation results from the Corr object.
<code>getUniquePairs</code>	Extract unique pairs from the Corr object.
<code>plot</code>	Plot the density distributions for correlation object(s).

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$. \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
corr<-Corr(examples$corrData,method="spearman",verbose=TRUE);
class(corr);
corr[1:10,];
```

CorrData

CorrData class

Description

Package: IdMappingAnalysis

Class CorrData**Object**

~~|

~~+--CorrData

Directly known subclasses:

public static class **CorrData**

extends [Object](#)

CorrData object stores the pair of experiments on which the correlation related processing is performed (MS/MS and mRNA for example) in such a way that two experiments are aligned by experiment names and by the primary keys ensuring the fast correlations. Typically, the primary ID of the ID Map set under consideration is a primary key for a first experiment, and the secondary ID if the ID Map set is a primary key for a second experiment. The alignment of two experiments by primary keys is guaranteed by using the unique pairs object to produce a matching pair of primary keys on which both experiments are ordered. Represented by a list of two elements with names corresponding to the primary and secondary IDs of the unique pairs ('acc' and 'probeset' for example), each element containing a data frame with primary or secondary IDs in the first column while the rest of columns contain the experiment data. The names of the data columns in both data frames are identical and correspond to the sample IDs. The match of sample IDs and an alignment by primary/secondary IDs is ensured by the proper processing during the object creation.

Usage

```
CorrData(uniquePairs=NULL, expSet1=NULL, expSet2=NULL, verbose=FALSE, ...)
```

Arguments

uniquePairs	UniquePairs object or a list of such objects on which a single or a list of CorrData objects is constructed.
expSet1	a first ExperimentSet object with primary IDs corresponding (partially intersecting) with the content the first column of UniquePairs (uniquePairsData) object.
expSet2	a second ExperimentSet object with primary IDs corresponding (partially intersecting) with the content the second column of UniquePairs (uniquePairsData) object.
verbose	if <code>TRUE</code> enables diagnostic messages. Default is <code>FALSE</code> .
...	Not used.

Fields and Methods**Methods:**

<code>as.MultiSet</code>	Convert CorrData object into MultiSet object.
<code>getExperimentSet</code>	Get experiment set data frame for a given modality.
<code>getSampleNames</code>	Get experiment sample names.
<code>getUniquePairs</code>	Extract unique pairs from the CorrData object.
<code>interactive.plot</code>	Draw a scatterplot of experiment data interactively.
<code>plot</code>	Scatterplot of experiment data.
<code>primaryKey</code>	Retrieves a primary key for a given CorrData object.
<code>secondaryKey</code>	Retrieves a secondary key for a given CorrData object.

Methods inherited from Object:

`$`, `$<-`, `[[`, `[[<-`, `as.character`, `attach`, `attachLocally`, `clearCache`, `clearLookupCache`, `clone`, `detach`, `equals`, `extend`, `finalize`, `gc`, `getEnvironment`, `getFieldModifier`, `getFieldModifiers`, `getFields`, `getInstantiationTime`, `getStaticInstance`, `hasField`, `hashCode`, `ll`, `load`, `objectSize`, `print`, `registerFinalizer`, `save`

Author(s)

Alex Lisovich, Roger Day

Examples

```
corrData<-CorrData(examples$uniquePairs,
examples$msmsExperimentSet,examples$mrnaExperimentSet,verbose=TRUE);
class(corrData);
```

DataFilter

*The DataFilter class***Description**

Package: IdMappingAnalysis

Class DataFilter[Object](#)

~~|

~~+--DataFilter

Directly known subclasses:public static class **DataFilter**extends [Object](#)

Serves as a wrapper for data data filtering functions define as static methods of the DataFilter class.

Usage

DataFilter()

Fields and Methods**Methods:**

do.apply	Filter experiment using constraints.
fisherTransform	Compute the Fisher transform.
fisherTransformInverse	Compute the Fisher inversed transform.
fisherTransformJacobean	Compute the Fisher transform Jacobean.
logTen	Compute log10 of a numerical vector combined with thresholding on minimum value.
minAvgCountConstraint	Perform mean based thresholding of an input vector.
minCountConstraint	Perform minimum count based thresholding of an input vector.
minCountGroupConstraint	Perform minimum count based thresholding of an input vector subdivided into groups.
removeNASeries	Remove NA series from the experiment set.

Methods inherited from Object:

\$, \$<-, [], [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Display

The Display class

Description

Package: IdMappingAnalysis

Class Display

[Object](#)

~~|

~~+--Display

Directly known subclasses:

public static class **Display**

extends [Object](#)

Serves as a wrapper for a set of graphical functions defined as static methods of Display class

Usage

Display()

Fields and Methods**Methods:**

copy	Save current plot to the file.
create	Open a new display device.
line.loess	Plot loess transformed data.
line.unsorted	Draw a curve from unsorted points.
progressMsg	Display a progress message.
textBoundingBox	Determine the size of the text bounding box.
zoom.pars	Zoom graphics parameters.

Methods inherited from Object:

\$, \$<-, [], [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

examples

IdMappingAnalysis sample data...

Description

IdMappingAnalysis sample data

Value

A list containing the following sample items:

mrnaExperimentSet	MRNA Experiment sample data frame
msmsExperimentSet	MSMS Experiment sample data frame
outcomeMap	outcome sample data frame
identDfList	list of ID Map sample data frames collected from various services
jointIdMap	sample JointIdMap object for exploring the various DBs quantitative identifier mapping performance
jointIdMap_corr	sample JointIdMap object for exploring the various DBs correlation based identifier mapping performance
uniquePairs	sample UniquePairs object
jointUniquePairs	sample JointUniquePairs object
corrData	sample CorrData object suitable for subsequent fast correlations
corr	sample Corr object encapsulating the correlation results
mixture	sample Mixture object encapsulating the mixture modeling results
bootstrap	sample Bootstrap object encapsulating the results of bootstrapping on correlations

Author(s)

Alex Lisovich, Roger Day

expectedUtility	<i>Expected utility of an ID mapping, ID filtering, or other bioinformatics data preparation method</i>
-----------------	---

Description

expectedUtility calculates mean expected utility and total expected utility across pairs of features from two bioinformatics platforms. It is used to evaluate an ID mapping, ID filtering, or other bioinformatics data preparation method.

Usage

```
expectedUtility(dataset, label = "",
  bootModelCorClusters,
  columnsToRemove = c("Utp", "Lfp", "deltaPlus", "pi1Hat"),
  Utp, Lfp, deltaPlus, guarantee = 1e-09)
```

Arguments

dataset	A data frame or list from a call to fit2clusters, the posterior probabilities for each observation, their variance estimates. See Details.
label	A text string describing the method being studied, to label the return value. This is handy for using rbind to combine results for different methods.
bootModelCorClusters	Source for mixture model estimates. If missing, extracted from calling frame.
columnsToRemove	Names of columns to remove from return value.
Utp	Utility of a true positive.
Lfp	Loss of a false positive.
deltaPlus	Parameter defined as $\Pr("+ "+" \text{ or } "0")$
guarantee	Minimum value for posterior probability.

Details

The input dataset should be a dataframe with one row per ID pair, and the following columns:

- Utp Utility of a true positive.
- Lfp Loss of a false positive.
- postProb The posterior probabilities for each observation
- postProbVar The variances of the posterior probabilities, usually estimated from the bootstrap using Boot

Value

A data frame with just one row. The columns are:

Utp	Utility of a true positive.
Lfp	Loss of a false positive.
deltaPlus	Parameter defined as $\Pr("+" "+" \text{ or } "0")$
deltaZero	Parameter defined as $\Pr("0" "0" \text{ or } "x")$
nPairs	Number of ID pairs selected by the method.
pi1Hat	The estimate of the probability of the high-correlation component; obtained from
PrPlus	Estimated probability that an ID pair is in the "+" group.
PrTrue	Estimated probability that an ID pair is in the "+" or "0" group: $\text{PrPlus}/\text{deltaPlus}$
PrFalse	Estimated probability that an ID pair is in the "-" group.
Utrue	The component of expected utility from "true positives": $\text{PrTrue} * \text{Utp}$.
Lfalse	The (negative) component of expected utility from "false positives": $\text{PrFalse} * \text{Lfp}$.
Eutility1	The average expected utility per ID pair: $\text{Utrue} - \text{Lfalse}$.
Eutility	The total expected utility, summing over ID pairs: $\text{nrow}(\text{dataset}) * \text{Eutility1}$.

fit2clusters

Flexible two-cluster mixture fit of a numeric vector

Description

fit2clusters uses an ECM algorithm to fit a two-component mixture model. It is more flexible than mclust in some ways, but it only deals with one-dimensional data.

Usage

```
fit2clusters(Y, Ylabel = "correlation", YsigSq,
  piStart = c(0.5, 0.5), VStart = c(0.1, 0.1),
  psiStart = c(0, 0.1), NinnerLoop = 1, nReps = 500,
  psi0Constraint, V0Constraint, sameV = FALSE,
  estimatesOnly = TRUE, plotMe = TRUE, testMe = FALSE,
  Ntest = 5000, simPsi = c(0, 0.4), simPi = c(2/3, 1/3),
  simV = c(0.05^2, 0.05^2), simAlpha = 5, simBeta = 400,
  seed, ...)
```

Arguments

Y	The vector of numbers to fit.
Ysigsq	The vector of variance estimates for Y.
Ylabel	Label for the Y axis in a density fit figure.
piStart	Starting values for the component proportions.
VStart	Starting values for the component variances.
psiStart	Starting values for the component means
NinnerLoop	Number of iterations in the "C" loop of ECM.
nReps	Upper limit of number of EM steps.
psi0Constraint	If not missing, a fixed value for the first component mean.
V0Constraint	If not missing, a fixed value for the first component variance.
sameV	If TRUE, the components have the same variance.
estimatesOnly	If TRUE, return only the estimates. Otherwise, returns details per observations, and return the estimates as an attribute.
plotMe	If TRUE, plot the mixture density and kernel smooth estimates.
testMe	If TRUE, run a code test.
Ntest	For testing purposes, the number of replications of simulated data.
simPsi	For testing purposes, the true means.
simPi	For testing purposes, the true proportions
simV	For testing purposes, the true variances.
simAlpha	For testing purposes, alpha parameter in rgamma for measurement error variance.
simBeta	For testing purposes, beta parameter in rgamma for measurement error variance.
seed	For testing purposes, random seed.
...	Not used; testing roxygen2.

Details

See the document "ECM_algorithm_for_two_clusters.pdf".

Value

If `estimatesOnly` is TRUE, return only the estimates: Otherwise, return a dataframe of details per observations, and return the `estimates` as an attribute. The estimates details are:

pi1	The probability of the 2nd mixture component
psi0	The mean of the first component (psi0Constraint if provided)
psi1	The mean of the second component
Var0	The variance of the first component (V0Constraint if provided)
Var1	The variance of the second component

The observations details are:

Y	The original observations.
Ysigsq	The original measurement variances.
posteriorOdds	Posterior odds of being in component 2 of the mixture.
postProbVar	Estimated variance of the posterior probability, using the delta method.

IdMap

The ID Map class

Description

Package: IdMappingAnalysis

Class IdMap

Object

```

~~|
~~+--IdMapBase
~~~~~|
~~~~~+--IdMap

```

Directly known subclasses:

```

public static class IdMap
extends IdMapBase

```

IdMap is an object encapsulating a data frame with two columns (Primary ID and Secondary ID) where primaryID is a character string uniquely identifying the ID under consideration (unprot accessions ID or acc, Entrez Gene ID etc) and the Secondary ID is a comma separated list of secondary IDs associated with a given primary ID for a particular DB service. The analysis typically starts from obtaining a set of ID Maps (from the various DB services) which are not assumed to have the same number of rows or the same set of primary IDs. The process of alignment of this ID Maps is performed within the JointIdMap

Usage

```
IdMap(DF=NULL, name="", primaryKey=colnames(DF)[1], secondaryKey=colnames(DF)[2], ...)
```

Arguments

DF	A data.frame consisting of two columns (primary and secondary IDs) from which the IdMap object is to be created.
name	A character string representing the name of the given IdMap object. Default is ""

primaryKey	The name of the primary (first) column in an ID Map. If missing then the input data frame first column name is used and if it is not available defaults to 'From'.
secondaryKey	The name of secondary (second) column in an ID Map. If missing then the input data frame second column name is used and if it is not available defaults to 'To'.
...	Not used.

Fields and Methods

Methods:

as	-
as.UniquePairs	Create a UniquePairs object from a given IdMap object.
as.list	Coerce an object or a list of compatible object .
getCounts	Compute the count of secondaryIDs for each primary ID.
merge	Merge the IdMap object with a second IdMap object or a list of IdMap objects.
swapKeys	Swap the primary and secondary key columns.

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$. \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
obj<-IdMap(examples$identDfList[[2]]);
obj$primaryKey();
obj$secondaryKey();
```

IdMapBase

The ID Map base class

Description

Package: IdMappingAnalysis

Class IdMapBase

[Object](#)

~~|

```
~~+--IdMapBase
```

Directly known subclasses:

[Bootstrap](#), [Corr](#), [IdMap](#), [IdMapCounts](#), [IdMapDiff](#), [IdMapDiffCounts](#), [JointIdMap](#), [JointUniquePairs](#), [UniquePairs](#)

```
public static class IdMapBase
extends Object
```

IdMapBase is an abstract object encapsulating a data frame with at least two columns, the first one (primary) containing character string s identifying the ID under consideration (unprot accessions ID or acc, Entrez Gene ID etc) and the rest of columns containing the various information associated with a given primary ID for a particular DB service.

Usage

```
IdMapBase(DF=NULL, name="", primaryKey=NULL, secondaryKey=NULL, ...)
```

Arguments

DF	A data.frame consisting of two columns (primary and secondary IDs) from which the IdMap object is to be created.
name	A character string representing the name of the given IdMap object. Default is ""
primaryKey	The primary identifier type from which the ID conversion is performed. If NULL (default) then the input data frame first column name is used and if it is not available defaults to 'From'.
secondaryKey	The secondary identifier type to which conversion is performed. Default is NULL .
...	Not used.

Fields and Methods

Methods:

[-
aligned	Checks if two IdMapBase objects match on column names and primary ID set.
as.data.frame	Retrieves a data frame encapsulated within the given IdMapBase object.
dim	Retrieves dimensions of data frame encapsulated within the given IdMapBase object.
dimnames	Retrieve or set the dimnames of data frame encapsulated within the given IdMapBase object.
getName	Get the name a given IdMapBase object.
primaryIDs	Retrieves the primary IDs for a given IdMapBase object.
primaryKey	Retrieves a primary key for a given IdMapBase object.
secondaryKey	Retrieves a secondary key for a given IdMapBase object.

Methods inherited from Object:

\$, \$<-, [, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
DF<-array(0,dim=c(5,2));
obj<-IdMapBase(DF,primaryKey="primary",secondaryKey="secondary");
```

IdMapCounts

The IdMapCounts class

Description

Package: IdMappingAnalysis

Class IdMapCounts

Object

```
~~|
~~+--IdMapBase
~~~~~|
~~~~~+--IdMapCounts
```

Directly known subclasses:

```
public static class IdMapCounts
  extends IdMapBase
```

An IdMapCounts object encapsulates a `data.frame` where the first column contains the primary ID set while the rest of columns contain the counts of secondary IDs for each Id Map in a given idMapList object, one column per ID Map, each ID Map related column having a name representing the given DB data source (i.e. 'NetAffx', 'EnVision' etc.) The constructor creates the IdMapCounts object from the list of ID Maps aligned by the primary IDs and primary and secondary keys. The easiest way to obtain the list of properly aligned IdMap objects is to create a JointIdMap object from a set of un-aligned ID maps and then invoke the getIdMapList() method on this object. The IdMapCounts object can also be created directly from JointIdMap object by using the JointIdMap.\$getCounts() method.

Usage

```
IdMapCounts(idMapList=NULL, verbose=FALSE, ...)
```


Arguments

idMapList	The list of ID Maps aligned on primary IDs.
verbose	If TRUE enables diagnostic messages. Default is FALSE
...	Not used.

Fields and Methods**Methods:**

getStats	Retrieves a set of unique counts of secondary IDs.
plot	Compute and plot the (inversed) ecdf for each ID Map count entry within the IdMapCounts object.

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$, \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
idMaps<-IdMap$as.list(examples$identDfList[[1]]);
cnts<-IdMapCounts(IdMap(examples$identDfList[[1]]));
cnts[1:20,];

#create IdMapCounts object from aligned IdMap list.
jointIdMap<-JointIdMap(examples$identDfList);
idMaps<-jointIdMap$getIdMapList(verbose=TRUE);
cnts<-IdMapCounts(idMaps);
cnts[1:20,];

#create IdMapCounts object directly from the JointIdMap object
jointIdMap<-JointIdMap(examples$identDfList);
cnts<-jointIdMap$getCounts(verbose=TRUE);
```

Description

Package: IdMappingAnalysis

Class IdMapDiff**Object**

```

~~|
~~+---IdMapBase
~~~~~|
~~~~~+---IdMapDiff

```

Directly known subclasses:

```

public static class IdMapDiff
extends IdMapBase

```

IdMapDiff constructor implements most time consuming step in comparing two DBs and the structure itself stores the results in a compact form. The IdMapDiff object encapsulates a [data.frame](#) the first column of which contains the primary IDs and the rest of columns contain a disjoint representation of the ID Map pair in the form of 3 columns <A-A*B,A*B,B-A*B>, where A and B are secondary ID lists for ID Maps A and B. This class is separated from the IdMapDiffCounts in anticipation of being used by various processing pipelines in a future.

Usage

```
IdMapDiff(idMap1=NULL, idMap2=NULL, pairNames=c("First", "Second"), verbose=FALSE, ...)
```

Arguments

idMap1	The first ID Map object on which IdMapDiff object is constructed.
idMap2	The second ID Map object on which IdMapDiff object is constructed.
pairNames	The character vector of length 2 representing the names of the ID Map pair. Default is c('First','Second').
verbose	If TRUE enables diagnostic messages.
...	Not used.

Fields and Methods**Methods:**

No methods defined.

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$. \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
#get primary IDs from an msms experiment set
IDs<-IdMapBase$primaryIDs(examples$msmsExperimentSet);

#create JointIdMap object aligned by primaryIDs
jointIdMap<-JointIdMap(examples$identDfList,primaryIDs=IDs);

# get IdMap list aligned of two ID maps aligned by primaryIDs
idMaps<-jointIdMap$getIdMapList(verbose=TRUE);

#create IdMapDiff object
diffs<-IdMapDiff(idMaps[["NetAffx_F"]],idMaps[["DAVID_Q"]]);
diffs[1:10,];

# create IdMapDiff object directly from JointIdMap
diffs<-jointIdMap$getDiff("NetAffx_F","DAVID_Q",verbose=TRUE);
```

IdMapDiffCounts

*The IdMapDiffCounts class***Description**

Package: IdMappingAnalysis

Class IdMapDiffCounts**Object**

```
~~|
~~+--IdMapBase
~~~~~|
~~~~~+--IdMapDiffCounts
```

Directly known subclasses:

```
public static class IdMapDiffCounts
extends IdMapBase
```

The IdMapDiffCounts class handles statistics on IdMapDiff object. IdMapDiffCounts object encapsulates a data frame with row names corresponding to the primary IDs and 6 columns subdivided into pairs <match(TRUE/FALSE),count> each pair corresponding to the disjoint events <A-A*B,A*B,B-A*B>, where A and B are secondary ID lists for ID Maps A and B from the IdMapDiff object. The 'pairNames' attribute of the IdMapDiffCounts contains the names of the source ID Map pair from which the IdMapDiff object was created.

Usage

```
IdMapDiffCounts(idMapDiff=NULL, verbose=FALSE, ...)
```

Arguments

idMapDiff	The IdMapDiff on which IdMapDiffCounts is created. Default is <code>NULL</code> .
verbose	If <code>TRUE</code> enables diagnostic messages. Default is <code>FALSE</code> .
...	Not used.

Fields and Methods**Methods:**

<code>getCompoundEvents</code>	Get compound events.
<code>getCompoundGroups</code>	Get counts for each compound event in IdMapDiffCounts.
<code>plot</code>	Produce a fountain plot representing the quantitative relationship of the compound events.
<code>summary</code>	Get a compound event counts summary report.

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$, \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
#get primary IDs from an msms experiment set
IDs<-IdMapBase$primaryIDs(examples$msmsExperimentSet);

#create JointIdMap object aligned by primaryIDs
jointIdMap<-JointIdMap(examples$identDfList,primaryIDs=IDs);

#create IdMapDiff object
diffs<-jointIdMap$getDiff("NetAffx_F", "DAVID_Q", verbose=TRUE);

# create IdMapDiffCounts object
diffCounts<-IdMapDiffCounts(diffs);
diffCounts[1:10,];
```

JointIdMap

*The Joint ID Map class***Description**

Package: IdMappingAnalysis

Class JointIdMap

Object

~~|

~~+---IdMapBase

~~~~~|

~~~~~+---JointIdMap

Directly known subclasses:public static class **JointIdMap**extends [IdMapBase](#)

JointIdMap is an object encapsulating a [data.frame](#) containing the primary ID set in a first column while the rest of columns containing the sets of secondary IDs, each column corresponding to a particular Id Map, keeping all Id Maps properly aligned

Usage

```
JointIdMap(idMapList=list(), primaryIDs=NULL, name="", verbose=FALSE, ...)
```

Arguments

| | |
|------------|--|
| idMapList | The list of ID Maps on which the JointData is constructed. |
| primaryIDs | The optional character vector of primary IDs on which an additional intersection and reordering are performed. |
| name | The optional name of a given JointIdMap object. Default is "" |
| verbose | if TRUE enables diagnostic messages. Default is @FALSE . |
| ... | Not used |

Fields and Methods**Methods:**

| | |
|---------------------------------|--|
| as.data.frame | Retrieve a data frame encapsulated within the given JointIdMap object. |
| diffCounts.plot | Interactive wrapper for IdMapDiffCounts\$plot. |
| ecdf.plot | Interactive wrapper for IdMapCounts\$plot. |

| | |
|----------------------------|---|
| <code>getCounts</code> | Create an IdMapCounts object. |
| <code>getDiff</code> | Create an IdMapDiff object. |
| <code>getIdMapList</code> | Create an Id Map list from a JointIdMap object. |
| <code>getMapNames</code> | Get the names of IdMap objects encapsulated within the given JointIdMap object. |
| <code>getMatchInfo</code> | Get match table(s) for a given set of primary IDs. |
| <code>getUnionIdMap</code> | Create a union IdMap. |

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$, \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Roger Day,Alex Lisovich

Examples

```
jointIdMap<-JointIdMap(examples$identDfList);

jointIdMap$primaryKey();
jointIdMap$secondaryKey();

jointIdMap[1:10,];
```

JointUniquePairs *The JointUniquePairs class*

Description

Package: IdMappingAnalysis

Class JointUniquePairs**Object**

```
~~|
~~+--IdMapBase
~~~~~|
~~~~~+--JointUniquePairs
```

Directly known subclasses:

```
public static class JointUniquePairs
  extends IdMapBase
```

UniquePairsMatch object encapsulates a data frame the first two columns of which contain the unique pairs corresponding to the merge (union) of all ID Maps in consideration while the rest of columns contains the match (logical value) between the merged unique pairs set and a unique pairs set specific to the particular ID Map ('d8', 'enV', 'netAffx' etc), one column per Id Map. Used in combination with correlation related data objects (CorrData, Corr, Mixture etc.) to aid in on-fly processing related to some classification by a particular match group The UniquePairsMatch constructor creates an object from the UniquePairs and an ID Map list computing the match (inclusions) for each particular ID Map.

Usage

```
JointUniquePairs(uniquePairs=NULL, idMapList=NULL, name="", verbose=FALSE, ...)
```

Arguments

| | |
|-------------|--|
| uniquePairs | UniquePairs object on which a UniquePairsMatch is created or a data.frame complying with the UniquePairs class internal data frame format. In case the UniquePairs object is used as a first argument, it's typically obtained from the JointIdMap object by invoking <code>JointIdMap\$getUnionIdMap()</code> |
| idMapList | the list of ID Maps on which the match is performed during the UniquePairs-Match object creation The idMapList typically obtained through the call to the <code>JointIdMap.getIdMapList()</code> of the same JointIdMap object as for the first argument to ensure that both arguments are properly aligned. |
| name | A character string representing the name of the given IdMap object. Default is "" |
| verbose | If TRUE enables diagnostic messages. Default is FALSE . |
| ... | Not used. |

Fields and Methods

Methods:

| | |
|----------------------------------|---|
| boxplot | Draw a basic boxplot based on a given JointUniquePairs object and external data. |
| corr.boxplot | Boxplot of correlations by match group. |
| corr.plot | Plot the density distributions for a set of correlation objects . |
| do.glm | Compute linear regression for the given set of ID Maps. |
| getBootstrap | Create Bootstrap object from JointUniquePairs object and two experiment sets. |
| getCorr | Extract a set of correlation objects from given JointUniquePairs object . |
| getCorrData | Create CorrData object from the JointUniquePairs object and two experiment sets. |
| getCorrDataFrame | Merge JointUniquePairs and Corr objects into a single data frame. |
| getMapNames | Get the names of UniquePairs objects encapsulated within the given JointUniquePairs |
| getMatchInfo | Get match table(s) for a given set of primary IDs. |
| getMixture | Extract mixture model object from JointUniquePairs and Corr objects. |
| getUniquePairs | Extract the unity UniquePairs object from a given JointUniquePairs object. |

| | |
|---|---|
| interactive.corr.boxplot | Interactive boxplot of correlations by match group. |
| interactive.corr.plot | Interactive plot of correlation densities. |
| interactive.mixture.boxplot | Interactive boxplot of mixture component probabilities by match group. |
| interactive.mixture.plot | Interactive plot of mixture model components. |
| interactive.plot | General purpose JointUniquePairs interactive plot function. |
| mixture.boxplot | Boxplot of a mixture model component by match group. |
| mixture.plot | Plot the correlation densities of the empirical fit, mixture fit and each . |
| subsetCorr | Subset the Corr object. |
| subsetData | Subset data on a UniquePairsMatch object. |
| subsetGroups | Get a JointUniquePairs subset. |

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$, \$<-, [, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
#create JointIdMap
jointIdMap<-JointIdMap(examples$identDfList);

#creaate unique pairs from the union of all IdMaps within JointIdMap
pairs<-as.UniquePairs(jointIdMap$getUnionIdMap(verbose=TRUE),verbose=TRUE);

#create JointUniquePairs object
jointPairs<-JointUniquePairs(pairs,jointIdMap$getIdMapList(),verbose=TRUE);
jointPairs[1:10,];
```

Misc

The Misc class

Description

Package: IdMappingAnalysis

Class Misc

[Object](#)

~~|


```
~+--Misc
```

Directly known subclasses:

```
public static class Misc
extends Object
```

Serves as a wrapper for various miscellaneous functions used throughout the package defined as static methods of the Misc class.

Usage

```
Misc()
```

Fields and Methods

Methods:

| | |
|--------------------------------|---|
| <code>CsvList.merge</code> | Pairwise merge of two string vectors. |
| <code>interleave</code> | Interleave two matrixes by columns. |
| <code>to.base</code> | Convert number to a numeric vector of a given base. |
| <code>to.binary.logical</code> | Convert number to a vector of logicals. |
| <code>to.index.expr</code> | Convert expression into index expression for a given list or data frame object. |
| <code>words</code> | Convert space delimited string to a vector of words. |

Methods inherited from Object:

`$`, `$<-`, `[]`, `[[<-`, `as.character`, `attach`, `attachLocally`, `clearCache`, `clearLookupCache`, `clone`, `detach`, `equals`, `extend`, `finalize`, `gc`, `getEnvironment`, `getFieldModifier`, `getFieldModifiers`, `getFields`, `getInstantiationTime`, `getStaticInstance`, `hasField`, `hashCode`, `ll`, `load`, `objectSize`, `print`, `registerFinalizer`, `save`

Author(s)

Alex Lisovich, Roger Day

Mixture

The Mixture class

Description

Package: IdMappingAnalysis
Class Mixture

```
Object
~~|
~~+--Mixture
```

Directly known subclasses:

```
public static class Mixture
extends Object
```

The constructor creates a model from a single `Corr` object using the number of clusters defined by `G` determining the optimal number of clusters by default and optionally using the Fisher transform.

Usage

```
Mixture(corr=NULL, G=c(1:5), Fisher=FALSE, verbose=FALSE, ...)
```

Arguments

| | |
|----------------------|--|
| <code>corr</code> | Corr object on which mixture modeling is performed. |
| <code>G</code> | number of components in mixture model. If <code>G</code> is a vector, the optimal number of components is determined. <code>G</code> is a vector (1:5) by default. |
| <code>Fisher</code> | if <code>TRUE</code> , the Fisher transform of correlation data is performed before the model is fitted. Default is <code>FALSE</code> . |
| <code>verbose</code> | if <code>TRUE</code> enables diagnostic messages. Default is <code>FALSE</code> . |
| <code>...</code> | Not used. |

Value

The resulting `Mixture` object encapsulates a data member `'model'` containing the results of mixture modeling represented by the `list` with following components:

| | |
|------------------------------|---|
| <code>corr</code> | the correlation data |
| <code>clust</code> | the clustering results data structure returned by <code>Mclust()</code> |
| <code>sd</code> | standard deviation derived from <code>clust\$parameters\$variance\$sigmasq</code> |
| <code>density</code> | the correlation density distribution |
| <code>marginalDensity</code> | the marginal density |

Fields and Methods

Methods:

| | |
|-----------------------|--|
| <code>clust</code> | Retrieve the clustering results data structure. |
| <code>getData</code> | Extract mixture component data from the <code>Mixture</code> object. |
| <code>getStats</code> | Get mixture component model summary info. |

[plot](#) Plot the results of mixture modeling.
[primaryKey](#) Retrieves a primary key for a given Mixture object.
[secondaryKey](#) Retrieves a secondary key for a given Mixture object.

Methods inherited from Object:

[\\$](#), [\\$<-](#), [\[\[](#), [\[\[<-](#), [as.character](#), [attach](#), [attachLocally](#), [clearCache](#), [clearLookupCache](#), [clone](#), [detach](#), [equals](#), [extend](#), [finalize](#), [gc](#), [getEnvironment](#), [getFieldModifier](#), [getFieldModifiers](#), [getFields](#), [getInstantiationTime](#), [getStaticInstance](#), [hasField](#), [hashCode](#), [ll](#), [load](#), [objectSize](#), [print](#), [registerFinalizer](#), [save](#)

Author(s)

Alex Lisovich, Roger Day

Examples

```

mixture<-Mixture(examples$corr,G=c(1:4),Fisher=TRUE,verbose=TRUE);
class(mixture);
names(mixture$.model)

```

Subset

The Subset class

Description

Package: IdMappingAnalysis

Class Subset

Object

```

~~|
~~+--Subset

```

Directly known subclasses:

```

public static class Subset
extends Object

```

Serves as a wrapper for data frame subsetting functions defined as static methods of the Subset class.

Usage

```
Subset(...)
```

Arguments

... Not used.

Fields and Methods**Methods:**

`byColNames` Extract subset of columns from a data frame or a list of data frames.
`byColumn` # Extract subset of rows from a data frame or a list of data frames .
`byRow` Extract subset of columns from a data frame or a list of data frames .
`byRowNames` Extract subset of columns from a data frame or a list of data frames.

Methods inherited from Object:

`$`, `$<-`, `[]`, `[[<-`, `as.character`, `attach`, `attachLocally`, `clearCache`, `clearLookupCache`, `clone`, `detach`, `equals`, `extend`, `finalize`, `gc`, `getEnvironment`, `getFieldModifier`, `getFieldModifiers`, `getFields`, `getInstantiationTime`, `getStaticInstance`, `hasField`, `hashCode`, `ll`, `load`, `objectSize`, `print`, `registerFinalizer`, `save`

Author(s)

Alex Lisovich, Roger Day

UniquePairs

The UniquePairs class

Description

Package: IdMappingAnalysis

Class UniquePairs

`Object`

~~|

~~+--IdMapBase

~~~~~|

~~~~~+--UniquePairs

Directly known subclasses:

public static class **UniquePairs**

extends `IdMapBase`

The alternative representation of an IdMap suitable for performing the correlation related processing. Contains a data frame with two columns, each row of which represents a unique pair <primary

ID, secondary ID> where primary ID corresponds to the primaryIDs of an ID Map and secondary ID corresponds to a single ID from a list of comma separated secondary IDs within the corresponding ID Map. The column names correspond to the primary/secondary keys of an Id Map ('acc' and 'probeset' for example)

Usage

```
UniquePairs(DF=NULL, name="", primaryKey=colnames(DF)[1], secondaryKey=colnames(DF)[2], ...)
```

Arguments

| | |
|--------------|---|
| DF | A <code>data.frame</code> consisting of two columns (primary and secondary IDs) from which the UniquePairs object is to be created. |
| name | A <code>character</code> string representing the name of the given UniquePairs object. Default is "" |
| primaryKey | The name of the primary (first) column of a <code>data.frame</code> encapsulated within the UniquePairs object. If missing then the input data frame first column name is used and if it is not available defaults to 'From'. |
| secondaryKey | The name of secondary (second) column in an ID Map. If missing then the input data frame second column name is used and if it is not available defaults to 'To'. |
| ... | Not used. |

Fields and Methods

Methods:

| | |
|-----------------------|---|
| <code>as.IdMap</code> | Convert the UniquePairs object into the IdMap object. |
| <code>as</code> | - |
| <code>create</code> | Create a UniquePairs object from a single IdMap or a list of IdMap objects. |
| <code>equals</code> | Check if two unique pairs data structures are identical. |
| <code>getMatch</code> | Get the logical vector of pair matches of the given UniquePairs object . |
| <code>swapKeys</code> | Swap the primary and secondary key columns. |
| <code>unique</code> | Extract unique elements. |

Methods inherited from IdMapBase:

[, aligned, as.data.frame, dim, dimnames, getName, primaryIDs, primaryKey, secondaryKey

Methods inherited from Object:

\$, \$<-, [[, [[<-, as.character, attach, attachLocally, clearCache, clearLookupCache, clone, detach, equals, extend, finalize, gc, getEnvironment, getFieldModifier, getFieldModifiers, getFields, getInstantiationTime, getStaticInstance, hasField, hashCode, ll, load, objectSize, print, registerFinalizer, save

Author(s)

Alex Lisovich, Roger Day

Examples

```
DF<-matrix(
  c("P25685", "200664_s_at",
    "P25685", "200666_s_at",
    "Q6ZV71", "205208_at",
    "Q6ZV71", "215798_at",
    "P05164", "203948_s_at"
  ), ncol=2, nrow=5, byrow=TRUE);
colnames(DF)<-c("Uniprot", "Affy");

uniquePairs<-UniquePairs(DF);
```

Index

*Topic **classes**

- Bootstrap, 2
- Corr, 4
- CorrData, 5
- DataFilter, 7
- Display, 8
- IdMap, 13
- IdMapBase, 14
- IdMapCounts, 16
- IdMapDiff, 17
- IdMapDiffCounts, 19
- JointIdMap, 21
- JointUniquePairs, 22
- Misc, 24
- Mixture, 25
- Subset, 27
- UniquePairs, 28

*Topic **data**

- examples, 9

- aligned, 15
- as.data.frame, 15, 21
- as.IdMap, 29
- as.list, 14
- as.MultiSet, 6
- as.UniquePairs, 14

- Bootstrap, 2, 15
- boxplot, 23
- byColNames, 28
- byColumn, 28
- byRow, 28
- byRowNames, 28

- character, 13, 15, 21, 23, 29
- clust, 26
- copy, 8
- Corr, 4, 15
- corr.boxplot, 23
- corr.plot, 23

- CorrData, 5
- create, 8, 29
- CsvList.merge, 25

- data.frame, 3, 4, 13, 15, 16, 18, 21, 23, 29
- DataFilter, 7
- diffCounts.plot, 21
- dim, 15
- dimnames, 15
- Display, 8
- do.apply, 7
- do.glm, 23

- ecdf.plot, 21
- equals, 29
- examples, 9
- expectedUtility, 10

- FALSE, 3, 4, 6, 17, 20, 23, 26
- fisherTransform, 7
- fisherTransformInverse, 7
- fisherTransformJacobian, 7
- fit2clusters, 11

- getBootstrap, 23
- getCompoundEvents, 20
- getCompoundGroups, 20
- getCorr, 23
- getCorrData, 23
- getCorrDataFrame, 23
- getCounts, 14, 22
- getData, 4, 26
- getDiff, 22
- getExperimentSet, 6
- getIdMapList, 22
- getMapNames, 22, 23
- getMatch, 29
- getMatchInfo, 22, 23
- getMixture, 23
- getName, 15

- getSampleNames, 6
- getStats, 17, 26
- getUnionIdMap, 22
- getUniquePairs, 4, 6, 23

- IdMap, 13, 15
- IdMapBase, 2, 4, 13, 14, 16, 18, 19, 21–23, 28
- IdMapCounts, 15, 16
- IdMapDiff, 15, 17
- IdMapDiffCounts, 15, 19
- interactive.corr.boxplot, 24
- interactive.corr.plot, 24
- interactive.mixture.boxplot, 24
- interactive.mixture.plot, 24
- interactive.plot, 6, 24
- interleave, 25

- JointIdMap, 15, 21
- JointUniquePairs, 15, 22

- line.loess, 8
- line.unsorted, 8
- list, 17, 21, 26
- logTen, 7

- merge, 14
- minAvgCountConstraint, 7
- minCountConstraint, 7
- minCountGroupConstraint, 7
- Misc, 24
- Mixture, 25
- mixture.boxplot, 24
- mixture.plot, 24

- NULL, 15, 20

- Object, 2, 4, 5, 7, 8, 13–16, 18, 19, 21, 22, 24–28

- plot, 3, 4, 6, 17, 20, 27
- primaryIDs, 15
- primaryKey, 6, 15, 27
- progressMsg, 8

- removeNASeries, 7

- secondaryKey, 6, 15, 27
- Subset, 27
- subsetCorr, 24
- subsetData, 24

- subsetGroups, 24
- summary, 20
- swapKeys, 14, 29

- textBoundingBox, 8
- to.base, 25
- to.binary.logical, 25
- to.index.expr, 25
- TRUE, 3, 4, 6, 17, 18, 20, 21, 23, 26

- unique, 29
- UniquePairs, 15, 28

- words, 25

- zoom.pars, 8