Package 'DRWPSurv'

May 3, 2016

Type Package

Title Predicting survival outcome using pathway activities

Version 1.0

Date 2016-05-02

Author Wei Liu

Maintainer Wei Liu <30330590@qq.com>

Depends igraph, Matrix, survival, glmnet

Description This package implements the DRWPSurv method which predicts survival outcome
using topologically inferred pathway activities.

License GPL(>=2)

NeedsCompilation no

R topics documented:

DRWPSurv-package   *Predicting survival outcome using pathway activities*

Description

This package implements the DRWPSurv method which predicts survival outcome using topologi-
cally inferred pathway activities.

Details

The DESCRIPTION file:

Package:	DRWPSurv
Type:	Package
Title:	Predicting survival outcome using pathway activities
Version:	1.0
Date:	2016-05-02
Author:	Wei Liu
Maintainer:	Wei Liu <30330590@qq.com>
Depends:	igraph, Matrix, survival, glmnet
Description:	This package implements the DRWPSurv method which predicts survival outcome using topologically infe
License:	GPL(>=2)

Index of help topics:

```
DRW                     Directed Random Walk
DRWPSurv-package        Predicting survival outcome using pathway
                        activities
dGMMirGraph             The global pathway graph
fit.DRWPSurv            Fit a Lasso-Cox model using DRWPSurv
getPathActivity         Inferring pathway activity
getW                    Calculating the weights of genes
mRNA_matrix             The expression data
pathSet                 Pathway set
predict.DRWPSurv        Make predictions from a "DRWPSurv" object
survData                Survival data
```

Very simple to use. Accepts x,y data for Lasso-Cox models, and makes predictions for new samples. Only 2 functions: `fit.DRWPSurv` topologically infers pathway activities and fits the Lasso-Cox model. `predict.DRWPSurv` predicts the risk of new patients using the Lasso-Cox model fitted by `fit.DRWPSurv`.

Author(s)

Wei Liu

Maintainer: Wei Liu <30330590@qq.com>

References

Liu, W., et al., Topologically inferring pathway activity for precise survival outcome prediction of breast cancer. Unpublished.

Examples

```
data(dGMMirGraph)
data(pathSet)
data(mRNA_matrix)
data(survData)
trainSmpl.Idx <- sample(1:dim(mRNA_matrix)[2], floor(4/5*dim(mRNA_matrix)[2]))
testSmpl.Idx <- setdiff(1:dim(mRNA_matrix)[2], trainSmpl.Idx)
trainSmpl <- mRNA_matrix[ ,trainSmpl.Idx]
testSmpl <- mRNA_matrix[ ,testSmpl.Idx]
fit <- fit.DRWPSurv(x.mRNA = t(trainSmpl), y = survData[trainSmpl.Idx,], DEBUG=TRUE,
                    standardize=TRUE, globalGraph = dGMMirGraph, pathSet = pathSet,
```

```
                        Gamma=0.7, alpha= 1, nfolds = 5)
predict.DRWPSurv(object = fit, newx.mRNA = t(testSmpl), type="link",s="lambda.min")
```

---

dGMMirGraph   *The global pathway graph*

Description

The directed global pathway graph constructed by the R package iSubpathwayMiner.

Usage

```
data("dGMMirGraph")
```

Format

An igraph R object.

Details

There are 7159 nodes and 39930 edges in dGMMirGraph. Each node in the graph represents a gene/miRNA/metabolite. The global pathway graph is used to evaluate the topological importance of genes by directed random walk.

Examples

```
data(dGMMirGraph)
```

DRW   *Directed Random Walk*

Description

The directed random walk algorithm proposed by Liu et al (2013).

Usage

```
DRW(igraphM, p0, EdgeWeight = FALSE, gamma = 0.3)
```

Arguments

igraphM	An igraph object containing the directed global pathway graph.
p0	A unit vector containing the initial weights of genes in the global pathway graph.
EdgeWeight	Logical. Should igraphM be converted to a weighted matrix or an un-weighted matrix (the default)?
gamma	A numeric value. The restart probability in directed random walk.

Details

This function implements the directed random walk algorithm proposed by Liu et al (2013). It evaluates the topological weight of each gene according to its topological importance in the global pathway graph. The genes that close to many other genes that have large initial weights will receive larger weights. The final weights reflect the topological importances of genes in the global pathway graph.

Value

A numerical vector containing the topological weights of nodes in `igraphM`.

Author(s)

Wei Liu <30330590@qq.com>

References

Liu, W., et al., Topologically inferring risk-active pathways toward precise cancer classification by directed random walk. Bioinformatics, 2013. 29(17): p. 2169-77.

Examples

```
data(dGMMirGraph)
vertexs <- V(dGMMirGraph)
p0 <- runif(length(vertexs), min = 0, max = 1)
names(p0) <- vertexs$name
p0 <- p0/sum(p0)
vertexWeight <- DRW(igraphM = dGMMirGraph, p0, EdgeWeight=FALSE, gamma = 0.3)
names(vertexWeight) <- names(p0)
```

---

fit.DRWPSurv   *Fit a Lasso-Cox model using DRWPSurv*

Description

DRWPSurv is a pathway-based survival prediction method which topologically infers survival associated pathway activities and constructs Lasso-Cox models using the pathway activities as features.

Usage

```
fit.DRWPSurv(x.mRNA, y, DEBUG = FALSE, standardize = TRUE, globalGraph, pathSet,
             Gamma = 0.7, alpha = 1, nfolds = 5)
```

Arguments

x.mRNA	a n x p matrix of gene expression measurements with n genes and p samples.
y	a p x 2 matrix of survival data. The two columns represent survival time 'time' and disease status 'status' respectively.
DEBUG	Logical. Should debugging information be plotted.
standardize	Logical flag for x.mRNA standardization, prior to fitting the model. Default is TRUE.

globalGraph	An igraph R object containing the global pathway graph.
pathSet	A list of pathways. Each pathway is represented as a vector of pathway member genes and metabolites.
Gamma	A numeric value. The restart probability in directed random walk. Default is Gamma = 0.7.
alpha	The elasticnet mixing parameter in glmnet.
nfolds	Number of folds - default is 10. Although nfolds can be as large as the sample size (leave-one-out CV), it is not recommended for large datasets. Smallest value allowable is nfolds=3.

Details

DRWPSurv integrates gene expression profiles and prior gene interaction information to topologically infer survival associated pathway activities, and uses the pathway activities as features to construct Lasso-Cox model. It uses topological importance of genes evaluated by directed random walk to enhance the robustness of pathway activities and thereby improve the predictive performance.

Value

Fitted "DRWPSurv" model object.

fit.cox	An object of class "cv.glmnet"
W	The weights of nodes in globalGraph.
geneCoxZP	The z statistic and p-value (univariate Cox regression) of nodes in globalGraph.
globalGraph	An igraph R object containing the global pathway graph.
pathSet	A list of pathways. Each pathway is represented as a vector of pathway member genes and metabolites.
features	The selected pathway features to construct the Lasso-Cox model.
sigGenes	A vector of genes used to infer the pathway activities in features.
sigPathGenes	A list of pathways in features. Each pathway contains the member genes used to infer its pathway activity.

Author(s)

Wei Liu

References

Liu, W., et al., Topologically inferring pathway activity for precise survival outcome prediction of breast cancer. Unpublished.

See Also

predict.DRWPSurv

Examples

```
data(dGMMirGraph)
data(pathSet)
data(mRNA_matrix)
data(survData)
trainSmpl.Idx <- sample(1:dim(mRNA_matrix)[2], floor(4/5*dim(mRNA_matrix)[2]))
# testSmpl.Idx <- setdiff(1:dim(mRNA_matrix)[2], trainSmpl.Idx)
trainSmpl <- mRNA_matrix[ ,trainSmpl.Idx]
# testSmpl <- mRNA_matrix[ ,testSmpl.Idx]
fit <- fit.DRWPSurv(x.mRNA = t(trainSmpl), y = survData[trainSmpl.Idx,], DEBUG=TRUE,
                    standardize=TRUE, globalGraph = dGMMirGraph, pathSet = pathSet,
                    Gamma=0.7, alpha= 1, nfolds = 5)
```

---

getPathActivity	*Inferring pathway activity*

Description

This function infers pathway activities using the DRWPSurv method.

Usage

```
getPathActivity(x, pathSet, w, vertexZP)
```

Arguments

x	A n x p matrix of gene expression measurements with n samples and p genes.
pathSet	A list of pathways. Each pathway is represented as a vector of pathway member genes and metabolites.
w	A numerical vector containing the topological weights of nodes in `globalGraph`.
vertexZP	A p x 2 matrix which contains the z statistic and p-value of p genes in `x`.

Details

For each pathway, we selected the genes with p-value < 0.05 (univariate cox regression) in the pathway to infer pathway activity. The expression values of genes are weighted by their topological weights obtained from directed random walk on the global pathway graph. Pathway activity inference transforms the gene expression profiles into pathway activity profiles, which are then used to fit the Lasso-Cox model.

Value

pathwayActivity	
	The pathway activities of pathways in `pathSet`.
sigGenes	The genes used to infer the pathway activity.

Author(s)

Wei Liu

References

Liu, W., et al., Topologically inferring risk-active pathways toward precise cancer classification by directed random walk. Bioinformatics, 2013. 29(17): p. 2169-77.

---

getW   *Calculating the weights of genes*

Description

This function gets the weights of nodes in `globalGraph`.

Usage

```
getW(geneWeight, globalGraph)
```

Arguments

geneWeight   The weights of genes.

globalGraph   The global pathway graph.

Value

The weights of nodes in `globalGraph`.

Author(s)

Wei Liu

---

mRNA_matrix   *The expression data*

Description

An example of breast cancer expression data with 2000 genes and 100 samples.

Usage

```
data("mRNA_matrix")
```

Format

The format is: num [1:2000, 1:100] 8.97 6.04 6.51 5.46 5.52 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:2000] "780" "5982" "3310" "7849" ... ..$: chr [1:100] "GSM79114" "GSM79115" "GSM79116" "GSM79118" ...

Examples

```
data(mRNA_matrix)
```

pathSet	*Pathway set*

Description

`pathSet` is composed of 306 KEGG pathways. Each pathway is represented as a vector of pathway member genes and metabolites.

Usage

```
data("pathSet")
```

Format

A list of 306 KEGG pathways.

Examples

```
data(pathSet)
```

predict.DRWPSurv	*Make predictions from a "DRWPSurv" object*

Description

This function predicts the risk of new samples from a fitted `DRWPSurv` object.

Usage

```
predict.DRWPSurv(object, newx.mRNA, type = "link", s = "lambda.min")
```

Arguments

object	Fitted `"DRWPSurv"` model object.
newx.mRNA	A matrix with new samples to predict.
type	Type of prediction required. Type "link" gives the linear predictors for "cox" models. See `predict.glmnet` for details.
s	Value(s) of the penalty parameter lambda at which predictions are required. See `predict.glmnet` for details.

Value

the fitted relative-risk of new patients in `newx.mRNA`.

Author(s)

Wei Liu

References

Liu, W., et al., Topologically inferring pathway activity for precise survival outcome prediction of breast cancer. Unpublished.

See Also

fit.DRWPSurv

Examples

```
data(dGMMirGraph)
data(pathSet)
data(mRNA_matrix)
data(survData)
trainSmpl.Idx <- sample(1:dim(mRNA_matrix)[2], floor(4/5*dim(mRNA_matrix)[2]))
testSmpl.Idx <- setdiff(1:dim(mRNA_matrix)[2], trainSmpl.Idx)
trainSmpl <- mRNA_matrix[ ,trainSmpl.Idx]
testSmpl <- mRNA_matrix[ ,testSmpl.Idx]
fit <- fit.DRWPSurv(x.mRNA = t(trainSmpl), y = survData[trainSmpl.Idx,], DEBUG=TRUE, standardize=TRUE, global
predict.DRWPSurv(object = fit, newx.mRNA = t(testSmpl), type="link",s="lambda.min")
```

---

survData	*Survival data*

Description

The survival data of patients in `mRNA_matrix`.

Usage

```
data("survData")
```

Format

The format is: num [1:100, 1:2] 0 0 0 0 0 1 0 0 1 1 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:100] "GSM79114" "GSM79115" "GSM79116" "GSM79118" ... ..$: chr [1:2] "status" "time"

Examples

```
data(survData)
```

Index