# Package 'longRPart'

January 2, 2012

**Type** Package

**Title** Recursive partitioning of longitudinal data using mixed-effects models

**Version** 1.0

**Date** 2008-03-7

**Author** Sam Stewart and Mohamed Abdolell

**Maintainer** Sam Stewart <samstewart11@gmail.com>

**Depends** nlme,rpart

**Description**

**License** GPL

**Repository** CRAN

**Date/Publication** 2009-01-08 07:57:00

## R topics documented:

`longRPart` *Recursive partitioning of longitudinal data*

## Description

This function uses the custom partitioning rules in **rpart** with the mixed effects models from **nlme** to construct regression trees out of longitudinal data.

## Usage

```
longRPart(lmeFormula,rPartFormula,randomFormula,data,weight=NULL,R=NULL,control = rpart.control())
```

## Arguments

| | |
|---|---|
| `lmeFormula` | The mixed-effects formula, as would be passed to the `lme` function |
| `rPartFormula` | a list of variables to be used as potential classifiers for the rpart function |
| `randomFormula` | The grouping equation, as would be passed to the `lme` function |
| `data` | The data set being analyzed |
| `weight` | A weighting of the observations to be passed to the rpart function. |
| `R` | The correlation matrix for the mixed effects model. Default is exponential, which is a generalization of the corAR1 structure |
| `control` | Standard rpart control structure |

## Details

Applying CART principles to longitudinal data using mixed effects models is outlined in the paper by Abdolell et al. refereneced below. The basic principle is the same as the more common partitioning processes, with the deviance function for longitudinal data being taken from the mixed effects model created at each node.

The arguments passed to the function can be classified as either rpart arguments or lme arguments: rPartFormula and control are used to build the rpart structure, while lmeFormula, randomFormula and R are used to build the lme models.

To pass data to the function, it must have one observation per row, with at least the following three columns: a response column, an id column (to connect observations on the same subject) and a time column to record the time of response. Other columns will be required for the covariates.

The function passes many of the arguments directly to the `rpart` or `lme` functions, so before the tree is a built, a regular mixed-effects model should be built using the `lme` function.

`rPartFormula` is a list of covariates that will be used as potential splits in `rpart`. The basic structure of the function is `~cov1+cov2+cov3+...`. Note that `~.` will not work in this situation, because rpart will try to include all the variables (including ID and tme) and the resulting classification will not make sense.

## Value

The object returned is a standard rpart object, with some additional attributes to be used in the plotting functions

## Author(s)

Mohamed Abdolell <mo@dal.ca> and Sam Stewart

## References

Abdolell et al. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. Statistics in Medicine (2002) vol. 21 pp. 3395-3409

## See Also

rpart lme corClasses lrpPlot lrpTreePlot lrpPVal lrpCI

## Examples

```
# data from Abdolell paper
data(pbkphData)
pbkphData$Time=as.factor(pbkphData$Time)

model = longRPart(pbkph~Time,~age+gender,~1|Subject,pbkphData,R=corExp(form=~time),control=rpart.control(minbuc
```

---

# lrpCI *Calculate confidence intervals for longitudinal regression trees*

---

## Description

Using bootstrapping to calculate confidence intervals for the first split of a longitudinal regression tree, as outlined in the referenced paper. The details are outlined in the linked pdf.

## Usage

```
lrpCI(model,B,alpha)
```

## Arguments

| | |
|---|---|
| model | model produced by longRPart |
| B | Number of bootstrap samples to run |
| alpha | The desired confidence level |

## Details

To calculate the confidence interval, B bootstrap samples of the patients are taken from the model data set, and a longitudinal regression tree is built, resulting B initial splitting values (one from each tree). The 100(1-a)% confidence interval is calculated using the quantiles of the B splitting values.

## Value

a vector of split values is returned, from which the confidence intervals can be obtained

## Author(s)

Mohamed Abdolell <mo@dal.ca> and Sam Stewart

## References

Abdolell et al. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. Statistics in Medicine (2002) vol. 21 pp. 3395-3409

## See Also

../doc/bootstrap.pdf

---

`lrpCV` *Performs cross-validation on the longRPart model*

---

## Description

This functions implements v-fold cross validation on the longRPart trees. IT IS CURRENTLY UNDER DEVELOPMENT AND SHOULD NOT BE USED FOR ANALYSIS.

## Usage

```
lrpCV(model)
```

## Arguments

| | |
|---|---|
| `model` | model produced by `longRPart` |

## Details

Using the xval in the the rpart control object, this function performs v-fold cross-validation. The object returned is deviance between the subject and the predicted values at the node it was navigated to as a test case. The mean of that vector would be the average deviances.

## Author(s)

Mohamed Abdolell <mo@dal.ca> and Sam Stewart

## References

Abdolell et al. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. Statistics in Medicine (2002) vol. 21 pp. 3395-3409

## lrpPlot *Line plot of the nodes of a longRPart object*

### Description

Plotting the results of a longRPart regression tree: each line in the plot represents a different node

### Usage

```
lrpPlot(model,smoothing="n",color=NULL,place="bottomright")
```

### Arguments

| | |
|---|---|
| model | Output from longRPart function |
| smoothing | Type of smoothing to apply to the lines (NOT IMPLEMENTED) |
| color | colors to label the lines, default is rainbow() |
| place | placement for the legend |

### Details

This function performs a simple line plot of the regression coefficients from the mixed-effects models of each of the nodes in the longitudinal regression tree. Eventually smoothing will be applied, but for now only 'n', for none, is implemented.

### Author(s)

Mohamed Abdolell <mo@dal.ca> and Sam Stewart

### See Also

longRPart lrpTreePlot

## lrpPVal *Calculate p-values for the longitudinal regression tree*

### Description

Using permutation tests, this function calculates the p-value for the first split of the tree; see the linked pdf for details. It also calculates the p-value for the entire tree, though the statistical theory behind this p-value is not as clear.

### Usage

```
lrpPVal(model,J)
```

### Arguments

| | |
|---|---|
| model | model produced by longRPart |
| J | number of permutations to run |

### Details

The basic principle behind the test is to randomized the covariates to the subject and then recalculate the tree: the p-value is the proportion of trees that have a greater reduction in deviance than the original tree.

### Note

J indicates that a longRPart tree will be built J times, and a reasonable value for J would be 100 (at least), which should take 100 times as long as the original longRPart function

### Author(s)

Mohamed Abdolell <mo@dal.ca> and Sam Stewart

### References

Abdolell et al. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. Statistics in Medicine (2002) vol. 21 pp. 3395-3409

### See Also

../doc/permutation.pdf

---

## lrpSubFunctions *Helper functions from the rpart library*

---

### Description

Helper functions used in the plotting process for longRPartTreePlot

### Usage

```
rpartco(tree, parms =  paste(".rpart.parms", dev.cur(), sep = "."))
tree.depth(nodes)
```

### Arguments

| | |
|---|---|
| tree | Model returned from rpart |
| parms | secondary data for the rpartco function |
| nodes | The nodes of a tree |

### Details

rpartco is used to find the (x,y) coordinates of a tree-plot.

tree.depth is used to find the depth of a set of nodes.

### Value

each return approriate vectors/lists

### See Also

lrpTreePlot plot.rpart

---

## lrpTreePlot *Standard rpart tree plot adapted to longitudinal data*

---

### Description

Using the standard **rpart** plotting function and subfunction rpartco, this function produces a tree plot with simple line plots at the nodes of the regression coefficients of a mixed effects model.

### Usage

```
lrpTreePlot(model,use.n=TRUE,colors=NULL,place='bottomright')
```

### Arguments

| | |
|---|---|
| model | Model produced by longRPart |
| use.n | Boolean indicating whether to include the number of observations at a leaf node in the plot. If F, then the slope of the line is printed instead |
| colors | The desired colors for each node. |
| place | THe placement of the legend |

### Details

This function uses the standard plot and text commands from **rpart** and longRPart, but includes lineplots of the mixed-effects models calculated at each node.

### Author(s)

Mohamed Abdolell <mo@dal.ca> and Sam Stewart

### See Also

longRPart lrpPlot

pbkphData *A dataset on the improvement in hearing over time for children with coclear implants*

## Description

This is the original dataset from Mohamed Abdolell's paper.

## Details

Note the format of this dataset, as this is the required format for the data to work. The data should be in LONG format, i.e., it should have one row per observation, which means that a patient should have as many rows as they have time points. For each row there needs to be a patient identifier, a time value and a response value. Also note that, for the variables that are being used as splits, they must be completely populated, and the same for each patient. For this particular dataset that means that age and gender are completely populated, having missing values in them would cause those observations to be dropped.

## References

Abdolell et al. Binary partitioning for continuous longitudinal data: categorizing a prognostic variable. Statistics in Medicine (2002) vol. 21 pp. 3395-3409

# Index