Package: longitree (via r-universe)

May 16, 2026

Title Decision Tree Analysis for Longitudinal Measurement Data

Version 1.0.1

Maintainer Ryoto Obata <`ryoto.obata@gmail.com`>

Description Implements tree-based methods for longitudinal data. The package constructs decision trees that evaluate both the main effect of a covariate and its interaction with time through a weighted splitting criterion. It supports single-tree construction, bootstrap-based multiple-tree selection, and tree visualisation. For methodological details, see Obata and Sugimoto (2026) <doi:10.1007/s11634-025-00665-2>.

License GPL (>= 3)

Encoding UTF-8

LazyData true

NeedsCompilation yes

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

Imports stats, graphics, partykit, ggparty, ggplot2, lme4, utils

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

Author Ryoto Obata [aut, cre], Tomoyuki Sugimoto [aut]

Config/pak/sysreqs cmake make

Repository https://ryt1207.r-universe.dev

Date/Publication 2026-05-16 17:28:41 UTC

RemoteUrl https://github.com/cran/longitree

RemoteRef HEAD

RemoteSha 7f54074c575885917689e05fdece4bc86e35a3f6

Contents

longitree 2
longitrees 5
ltreedata 7
selectionplot 7
threetrees 8
treeplot 9

Index **11**

`longitree` *Construction of a Decision Tree for Longitudinal Data*

Description

Constructs a single decision tree for longitudinal data. The method evaluates both the main effect of a covariate and its interaction with time, incorporating a weighting mechanism to balance the two effects. Three single-tree construction procedures (ST1, ST2, ST3) are available; see Details. For the underlying methodology, refer to Obata and Sugimoto (2026).

Usage

```
longitree(
  formula,
  time,
  random,
  weight = "w",
  data,
  alpha = "no",
  gamma = "no",
  cv = "no",
  maxdepth = 5,
  minbucket = 5,
  minsplit = 20,
  xval = 10
)

## S3 method for class 'longitree'
summary(object, ...)

## S3 method for class 'longitree'
print(x, ...)

## S3 method for class 'longitree'
predict(object, ...)

## S3 method for class 'longitree'
plot(x, ...)
```

Arguments

`formula`	A formula specifying the model. The response variable should be on the left side and covariates on the right side. Use `response ~ .` to include all covariates except the time variable and the random effect, or select specific covariates such as `response ~ x1 + x2`. Time-invariant (baseline) covariates are assumed.
`time`	Character string giving the column name of the time variable. All individuals are assumed to be observed at the same time points.
`random`	Character string giving the column name of the random effect (subject identifier).
`weight`	Weight for balancing the main effect of a covariate and its interaction with time. A value in $\{0.0, 0.1, \ldots, 1.0\}$: `1.0` evaluates only the mean difference in the response variable between the two groups and `0.0` evaluates only the difference in change over time of the response variable between the two groups. Set `weight = "w"` (the default) to select the optimal weight from the same grid at each node.
`data`	A data frame containing the variables in `formula` together with the time and random-effect variables.
`alpha`	Significance level used as the stopping rule for tree growth. A smaller value produces a more conservative (smaller) tree. Specify a numeric value or `"no"` (default) if not used. Corresponds to ST2.
`gamma`	Complexity parameter for pruning. A larger value prunes more aggressively, yielding a smaller and simpler tree; a smaller value retains more branches. Specify a numeric value or `"no"` (default) if not used. Corresponds to ST3.
`cv`	Set `"yes"` to construct the decision tree using cross-validation, or `"no"` (default) otherwise. Corresponds to ST1.
`maxdepth`	Maximum depth of the tree (default 5).
`minbucket`	Minimum number of subjects in a terminal node (default 5).
`minsplit`	Minimum number of subjects required to attempt a split (default 20).
`xval`	Number of cross-validation folds (default 10). Used to compute the cross-validated coefficient of determination (R^2_{CV}); when `cv = "yes"`, also used for final tree selection.
`object`	A `longitree` object.
`...`	Additional arguments passed to `treeplot`.
`x`	A `longitree` object.

Details

Exactly one of `alpha`, `gamma`, or `cv` must be specified. Specifying more than one will result in an error. These correspond to the three single-tree construction procedures:

- **ST1 (**`cv = "yes"`**)** Tree growth, pruning, and final tree selection via cross-validation.
- **ST2 (**`alpha`**)** Tree growth with a significance threshold. No pruning or final tree selection via cross-validation.
- **ST3 (**`gamma`**)** Tree growth followed by pruning with a pre-specified complexity parameter. No final tree selection via cross-validation.

Since the time variable is not used as a splitting variable, each terminal node (leaf) contains the full longitudinal responses for every subject assigned to it, allowing direct evaluation of longitudinal trajectories within each leaf.

Value

An object of class `"longitree"`. Use `summary.longitree`, `predict.longitree`, or `plot.longitree` to inspect the results.

Methods (by generic)

- `summary(longitree)`: Print a brief summary of a `longitree` object.
- `print(longitree)`: Print method (calls `summary`).
- `predict(longitree)`: Extract predicted values and terminal node assignments from a `longitree` object. Returns a data frame with columns `predict` (predicted values) and `terminalnode` (terminal node assignments).
- `plot(longitree)`: Plot a `longitree` object. A convenience wrapper around `treeplot`.

References

Obata, R. and Sugimoto, T. (2026). A decision tree analysis for longitudinal measurement data and its applications. *Advances in Data Analysis and Classification*. doi:10.1007/s11634025006652

See Also

`treeplot`, `longitrees`

Examples

```
data(ltreedata)
# ST1: tree construction via cross-validation
result_st1 <- longitree(y ~ ., time = "time", random = "subject",
                           weight = 0.7, data = ltreedata, cv = "yes")
summary(result_st1)
predict(result_st1)
plot(result_st1)

# ST2: tree growth with a significance threshold
result_st2 <- longitree(y ~ ., time = "time", random = "subject",
                           weight = 0.1, data = ltreedata, alpha = 0.05)
summary(result_st2)
predict(result_st2)
plot(result_st2)

# ST3: pruning with a complexity parameter
result_st3 <- longitree(y ~ ., time = "time", random = "subject",
                           weight = "w", data = ltreedata, gamma = 3)
summary(result_st3)
predict(result_st3)
plot(result_st3)
```

`longitrees` *Construction of Multiple Decision Trees for Longitudinal Data*

Description

Generates multiple trees from bootstrap samples and evaluates all three-tree combinations based on two criteria: cross-validated prediction error and tree diversification measured by the adjusted Rand index (ARI). Bootstrap sampling is performed at the subject level to preserve longitudinal structure.

Usage

```
longitrees(
  formula,
  time,
  random,
  weight = "w",
  data,
  alpha = "no",
  gamma = "no",
  cv = "no",
  maxdepth = 5,
  minbucket = 5,
  minsplit = 20,
  xval = 10,
  bootsize,
  trees = 100,
  mins = 40
)
```

Arguments

`formula`	A formula specifying the model. The response variable should be on the left side and covariates on the right side. Use `response ~ .` to include all covariates except the time variable and the random effect, or select specific covariates such as `response ~ x1 + x2`. Time-invariant (baseline) covariates are assumed.
`time`	Character string giving the column name of the time variable. All individuals are assumed to be observed at the same time points.
`random`	Character string giving the column name of the random effect (subject identifier).
`weight`	Weight for balancing the main effect of a covariate and its interaction with time. A value in $\{0.0, 0.1, \ldots, 1.0\}$: `1.0` evaluates only the mean difference in the response variable between the two groups and `0.0` evaluates only the difference in change over time of the response variable between the two groups. Set `weight = "w"` (the default) to select the optimal weight from the same grid at each node.
`data`	A data frame containing the variables in `formula` together with the time and random-effect variables.

`alpha`	Significance level used as the stopping rule for tree growth. A smaller value produces a more conservative (smaller) tree. Specify a numeric value or `"no"` (default) if not used. Corresponds to ST2.
`gamma`	Complexity parameter for pruning. A larger value prunes more aggressively, yielding a smaller and simpler tree; a smaller value retains more branches. Specify a numeric value or `"no"` (default) if not used. Corresponds to ST3.
`cv`	Set `"yes"` to construct the decision tree using cross-validation, or `"no"` (default) otherwise. Corresponds to ST1.
`maxdepth`	Maximum depth of the tree (default 5).
`minbucket`	Minimum number of subjects in a terminal node (default 5).
`minsplit`	Minimum number of subjects required to attempt a split (default 20).
`xval`	Number of cross-validation folds (default 10). Used to compute the cross-validated coefficient of determination (R^2_{CV}); when `cv = "yes"`, also used for final tree selection.
`bootsize`	Number of subjects in each bootstrap sample.
`trees`	Number of bootstrap trees to grow (default 100).
`mins`	Number of top-ranking candidate three-tree subsets to retain (default 40).

Details

See `longitree` for a description of the three single-tree construction procedures (ST1, ST2, ST3) corresponding to `cv`, `alpha`, and `gamma`.

Value

An object of class `"longitrees"`. Pass to `selectionplot` to select the optimal three-tree combination.

References

Obata, R. and Sugimoto, T. (2026). A decision tree analysis for longitudinal measurement data and its applications. *Advances in Data Analysis and Classification*. doi:10.1007/s11634025006652

See Also

`longitree`, `selectionplot`, `threetrees`, `treeplot`

`ltreedata` *Sample longitudinal data for decision tree examples*

Description

A sample balanced longitudinal dataset with 50 subjects observed at 10 equally spaced time points.

Usage

```
ltreedata
```

Format

A data frame with 500 rows and 7 variables:

y Response variable (continuous).

subject Subject identifier (integer, 1–50).

time Time point (integer, 1–10).

x1 Baseline covariate 1 (integer, 1–10).

x2 Baseline covariate 2 (integer, 1–10).

x3 Baseline covariate 3 (integer, 1–6).

x4 Baseline covariate 4 (integer, 1–12).

`selectionplot` *Select Optimal Three-Tree Combination*

Description

Plots the cross-validated prediction error against the maximum pairwise adjusted Rand index (ARI) for candidate three-tree subsets, and selects a subset based on either prediction performance or tree diversification. The selected combination is indicated by a red point on the plot, which corresponds to the three trees used in the subsequent `threetrees` step.

Usage

```
selectionplot(longitrees, metric, nth)
```

Arguments

`longitrees`	A `longitrees` object.
`metric`	`"PE"` to select the subset with the smallest cross-validated prediction error, or `"ARI"` to select the subset with the smallest maximum pairwise ARI (greatest tree diversification).
`nth`	Rank of the tree subset to select (1 = best).

Value

An object of class `"selectionplot"`. Pass to `threetrees` to refit and evaluate the selected trees.

References

Obata, R. and Sugimoto, T. (2026). A decision tree analysis for longitudinal measurement data and its applications. *Advances in Data Analysis and Classification*. doi:10.1007/s11634025006652

See Also

`longitrees`, `threetrees`

`threetrees` *Fit and Evaluate Three Selected Trees*

Description

Refits the three trees selected by `selectionplot` on their original bootstrap samples.

Usage

```
threetrees(x, selection)

## S3 method for class 'threetrees'
summary(object, ...)

## S3 method for class 'threetrees'
print(x, ...)

## S3 method for class 'threetrees'
predict(object, tree = 1, ...)

## S3 method for class 'threetrees'
plot(x, tree = 1, ...)
```

Arguments

`x`	A `threetrees` object.
`selection`	A `selectionplot` object.
`object`	A `threetrees` object.
`...`	Additional arguments passed to `treeplot`.
`tree`	Integer 1, 2, or 3 selecting which tree to plot.

Value

An object of class `"threetrees"`. Use `summary.threetrees`, `predict.threetrees`, or `plot.threetrees` to inspect the results.

Methods (by generic)

- `summary(threetrees)`: Print a brief summary of a `threetrees` object.
- `print(threetrees)`: Print method (calls `summary`).
- `predict(threetrees)`: Extract predicted values and terminal node assignments from a `threetrees` object. Returns a data frame with columns `predict` (predicted values) and `terminalnode` (terminal node assignments).
- `plot(threetrees)`: Plot one of the three trees. A convenience wrapper around `treeplot`.

References

Obata, R. and Sugimoto, T. (2026). A decision tree analysis for longitudinal measurement data and its applications. *Advances in Data Analysis and Classification*. doi:10.1007/s11634025006652

See Also

`longitrees`, `selectionplot`, `treeplot`

Examples

```
data(ltreedata)
set.seed(10)
trees_res <- longitrees(y ~ ., time = "time", random = "subject",
                           weight = 0.5, data = ltreedata, alpha = 0.01,
                           bootsize = 50, mins = 40)
sel <- selectionplot(trees_res, metric = "PE", nth = 1)
tt <- threetrees(trees_res, selection = sel)
summary(tt)
predict(tt, tree = 1)
predict(tt, tree = 2)
predict(tt, tree = 3)
plot(tt, tree = 1)
plot(tt, tree = 2)
plot(tt, tree = 3)
```

`treeplot` *Decision Tree Plot Visualisation for Longitudinal Data*

Description

Visualises the structure of a decision tree for longitudinal data. Built on **ggparty**. Each split node displays the node number, split variable, p-value, and weight w. Each terminal node displays the node number, sample size N, and the intercept ($\hat{\beta}_0$) and slope ($\hat{\beta}_1$) from a linear mixed-effects model fitted within that node. Individual longitudinal trajectories are shown as dashed lines; the predicted values (average at each time point) are shown as solid lines, with the response variable on the vertical axis and time on the horizontal axis.

Usage

```
treeplot(
  x,
  tree = NULL,
  snsize = 50,
  spsize = 5,
  plotsize = 80,
  linesize1 = 0.3,
  linesize2 = 1,
  tnsize = 60
)
```

Arguments

`x`	A `longitree` or `threetrees` object.
`tree`	Integer 1, 2, or 3 selecting which tree to plot when `x` is a `threetrees` object.
`snsize`	Split-node label size (default 50).
`spsize`	Split-point label size (default 5).
`plotsize`	Overall plot size (default 80).
`linesize1`	Branch line width (default 0.3).
`linesize2`	Main line width (default 1).
`tnsize`	Terminal-node label size (default 60).

Value

A `ggplot2/ggparty` object.

References

Obata, R. and Sugimoto, T. (2026). A decision tree analysis for longitudinal measurement data and its applications. *Advances in Data Analysis and Classification*. doi:10.1007/s11634025006652

See Also

`longitree`, `threetrees`

Index

∗ **datasets**
    `ltreedata`, 7

`longitree`, 2, *6*, *10*
`longitrees`, *4*, 5, *8*, *9*
`ltreedata`, 7

`plot.longitree`, *4*
`plot.longitree` *(*`longitree`*)*, 2
`plot.threetrees`, *8*
`plot.threetrees` *(*`threetrees`*)*, 8
`predict.longitree`, *4*
`predict.longitree` *(*`longitree`*)*, 2
`predict.threetrees`, *8*
`predict.threetrees` *(*`threetrees`*)*, 8
`print.longitree` *(*`longitree`*)*, 2
`print.threetrees` *(*`threetrees`*)*, 8

`selectionplot`, *6*, 7, *8*, *9*
`summary.longitree`, *4*
`summary.longitree` *(*`longitree`*)*, 2
`summary.threetrees`, *8*
`summary.threetrees` *(*`threetrees`*)*, 8

`threetrees`, *6*, *7*, 8, *8*, *10*
`treeplot`, *3*, *4*, *6*, *8*, 9, *9*