

Package ‘glmertree’

February 20, 2019

Title Generalized Linear Mixed Model Trees

Date 2019-02-20

Version 0.1-2

Description Recursive partitioning based on (generalized) linear mixed models (GLMMs) combining lmer()/glmer() from 'lme4' and lmtree()/glmertree() from 'partykit'. The fitting algorithm is described in more detail in Fokkema, Smits, Zeileis, Hothorn & Kelderman (2018; <DOI:10.3758/s13428-017-0971-x>).

Depends R (>= 3.1.0), lme4, partykit (>= 1.0-4)

Imports graphics, stats, utils, Formula

Suggests vcd, lattice

LazyData yes

License GPL-2 | GPL-3

RoxygenNote 6.1.0

NeedsCompilation no

Author Marjolein Fokkema [aut, cre],
Achim Zeileis [aut]

Maintainer Marjolein Fokkema <M.Fokkema@fsw.leidenuniv.nl>

Repository CRAN

Date/Publication 2019-02-20 22:40:03 UTC

R topics documented:

DepressionDemo	2
glmertree	3
MHserviceDemo	6
Index	8

DepressionDemo

Artificial depression treatment dataset

Description

Simulated dataset of a randomized clinical trial (N = 150) to illustrate fitting of (G)LMM trees.

Usage

```
data("DepressionDemo")
```

Format

A data frame containing 150 observations on 6 variables:

depression numeric. Continuous treatment outcome variable (range: 3-16, M = 9.12, SD = 2.66).

treatment factor. Binary treatment variable.

cluster factor. Indicator for cluster with 10 levels.

age numeric. Continuous partitioning variable (range: 18-69, M = 45, SD = 9.56).

anxiety numeric. Continuous partitioning variable (range: 3-18, M = 10.26, SD = 3.05).

duration numeric. Continuous partitioning variable (range: 1-17, M = 6.97, SD = 2.90).

depression_bin factor. Binarized treatment outcome variable (0 = recovered, 1 = not recovered).

Details

The data were generated such that the duration and anxiety covariates characterized three subgroups with differences in treatment effects. The cluster variable was used to introduce a random intercept that should be accounted for. The treatment outcome is an index of depressive symptomatology.

See Also

[lmertree](#), [glmertree](#)

Examples

```
data("DepressionDemo", package = "glmertree")
summary(DepressionDemo)
lt <- lmertree(depression ~ treatment | cluster | anxiety + duration + age,
              data = DepressionDemo)
plot(lt)
gt <- glmertree(depression_bin ~ treatment | cluster | anxiety + duration + age,
               data = DepressionDemo)
plot(gt)
```

glmtree

*(Generalized) Linear Mixed Model Trees***Description**

Model-based recursive partitioning based on (generalized) linear mixed models.

Usage

```
lmtree(formula, data, weights = NULL, cluster = NULL,
        ranefstart = NULL, offset = NULL, joint = TRUE,
        abstol = 0.001, maxit = 100, dfsplit = TRUE, verbose = FALSE,
        plot = FALSE, lmer.control = lmerControl(), ...)
```

```
glmtree(formula, data, family = "binomial", weights = NULL,
        cluster = NULL, ranefstart = NULL, offset = NULL, joint = TRUE,
        abstol = 0.001, maxit = 100, dfsplit = TRUE, verbose = FALSE,
        plot = FALSE, glmer.control = glmerControl(), ...)
```

Arguments

formula	formula specifying the response variable and a three-part right-hand-side describing the regressors, random effects, and partitioning variables, respectively. For details see below.
data	data.frame to be used for estimating the model tree.
family	family specification for <code>glm</code> and <code>glmer</code> . See glm documentation for families.
weights	numeric. An optional numeric vector of weights. Can be a name of a column in data or a vector of length <code>nrow(data)</code> .
cluster	optional vector of cluster IDs to be employed for clustered covariances in the parameter stability tests. Can be a name of a column in data or a vector of length <code>nrow(data)</code> . If <code>cluster = NULL</code> (the default), observation-level covariances are employed in the parameter stability tests. If partitioning variables are measured on the cluster level, this will likely yield spurious splits, which can be mitigated by specification of the cluster argument, which results in cluster-level covariances being employed in the parameter stability tests.
ranefstart	NULL (the default), TRUE, or a numeric vector of length <code>nrow(data)</code> . Specifies the offset to be used in estimation of the first tree. NULL by default, yielding a zero offset initialization. If <code>ranefstart = TRUE</code> is specified, the random effects will be estimated first and the first tree will be grown using the random-effects predictions as an offset.
offset	optional numeric vector to be included in the linear predictor with a coefficient of one. Note that <code>offset</code> can be a name of a column in data or a numeric vector of length <code>nrow(data)</code> .

<code>joint</code>	logical. Should the fixed effects from the tree be (re-)estimated jointly along with the random effects?
<code>abstol</code>	numeric. The convergence criterion used for estimation of the model. When the difference in log-likelihoods of the random-effects model from two consecutive iterations is smaller than <code>abstol</code> , estimation of the model tree has converged.
<code>maxit</code>	numeric. The maximum number of iterations to be performed in estimation of the model tree.
<code>dfsplitt</code>	logical or numeric. <code>as.integer(dfsplitt)</code> is the degrees of freedom per selected split employed when extracting the log-likelihood.
<code>verbose</code>	Should the log-likelihood value of the estimated random-effects model be printed for every iteration of the estimation?
<code>plot</code>	Should the tree be plotted at every iteration of the estimation? Note that selecting this option slows down execution of the function.
<code>lmer.control</code> , <code>glmer.control</code>	list. An optional list with control parameters to be passed to <code>lmer()</code> and <code>glmer()</code> , respectively. See lmerControl and glmerControl for details.
<code>...</code>	Additional arguments to be passed to <code>lmtree()</code> or <code>glmmtree()</code> . See mob_control documentation for details.

Details

(G)LMM trees learn a tree where each terminal node is associated with different fixed-effects regression coefficients while adjusting for global random effects (such as a random intercept). This allows for detection of subgroups with different fixed-effects parameter estimates, keeping the random effects constant throughout the tree (i.e., random effects are estimated globally). The estimation algorithm iterates between (1) estimation of the tree given an offset of random effects, and (2) estimation of the random effects given the tree structure. See Fokkema et al. (2018) for a detailed introduction.

To specify all variables in the model a formula such as `y ~ x1 + x2 | random | z1 + z2 + z3` is used, where `y` is the response, `x1` and `x2` are the regressors in every node of the tree, `random` is the random effects, and `z1` to `z3` are the partitioning variables considered for growing the tree. If `random` is only a single variable such as `id` a random intercept with respect to `id` is used. Alternatively, it may be an explicit random-effects formula such as `(1 | id)` or a more complicated formula such as `((1+time) | id)`. (Note that in the latter two formulas, the brackets are necessary to protect the pipes in the random-effects formulation.)

In the random-effects model from step (2), two strategies are available: Either the fitted values from the tree can be supplied as an offset (`joint = FALSE`) so that only the random effects are estimated. Or the fixed effects are (re-)estimated along with the random effects using a nesting factor with nodes from the tree (`joint = TRUE`). In the former case, the estimation of each random-effects model is typically faster, but more iterations are required.

The code is still under development and might change in future versions.

Value

The function returns a list with the following objects:

`tree` The final `lmtree/glmmtree`.

<code>lmer</code>	The final lmer random-effects model.
<code>ranef</code>	The corresponding random effects of lmer.
<code>varcorr</code>	The corresponding <code>VarCorr(lmer)</code> .
<code>variance</code>	The corresponding <code>attr(VarCorr(lmer), "sc")^2</code> .
<code>data</code>	The dataset specified with the <code>data</code> argument including added auxiliary variables <code>.ranef</code> and <code>.tree</code> from the last iteration.
<code>loglik</code>	The log-likelihood value of the last iteration.
<code>iterations</code>	The number of iterations used to estimate the lmer tree.
<code>maxit</code>	The maximum number of iterations specified with the <code>maxit</code> argument.
<code>ranefstart</code>	The random effects used as an offset, as specified with the <code>ranefstart</code> argument.
<code>formula</code>	The formula as specified with the <code>formula</code> argument.
<code>randomformula</code>	The formula as specified with the <code>randomformula</code> argument.
<code>abstol</code>	The prespecified value for the change in log-likelihood to evaluate convergence, as specified with the <code>abstol</code> argument.
<code>mob.control</code>	A list containing control parameters passed to <code>lmtree()</code> , as specified with <code>...</code>
<code>lmer.control</code>	A list containing control parameters passed to <code>lmer()</code> , as specified in the <code>lmer.control</code> argument.
<code>joint</code>	Whether the fixed effects from the tree were (re-)estimated jointly along with the random effects, specified with the <code>joint</code> argument.

References

Fokkema M, Smits N, Zeileis A, Hothorn T, Kelderman H (2018). "Detecting Treatment-Subgroup Interactions in Clustered Data with Generalized Linear Mixed-Effects Model Trees". *Behavior Research Methods*, 50(5), 2016-2034. <https://doi.org/10.3758/s13428-017-0971-x>

See Also

[lmer](#), [glmer](#), [lmtree](#), [glmtree](#)

Examples

```
## artificial example data
data("DepressionDemo", package = "glmertree")

## fit normal linear regression LMM tree for continuous outcome
lt <- lmertree(depression ~ treatment | cluster | age + anxiety + duration,
  data = DepressionDemo)
print(lt)
plot(lt, which = "all") # default behavior, may also be "tree" or "ranef"
coef(lt)
ranef(lt)
predict(lt, type = "response") # default behavior, may also be "node"
predict(lt, re.form = NA) # excludes random effects, see ?lme4::predict.merMod
residuals(lt)
```

```

VarCorr(lt) # see lme4::VarCorr

## fit logistic regression GLMM tree for binary outcome
gt <- glmerTree(depression_bin ~ treatment | cluster | age + anxiety + duration,
  data = DepressionDemo)
print(gt)
plot(gt, which = "all") # default behavior, may also be "tree" or "ranef"
coef(gt)
ranef(gt)
predict(gt, type = "response") # default behavior, may also be "node" or "link"
predict(gt, re.form = NA) # excludes random effects, see ?lme4::predict.merMod
residuals(gt)
VarCorr(gt) # see lme4::VarCorr

```

MHserviceDemo

Artificial mental-health service outcomes dataset

Description

Artificial dataset of treatment outcomes (N = 3739) of 13 mental-health services to illustrate fitting of (G)LMM trees with constant fits in terminal nodes.

Usage

```
data("MHserviceDemo")
```

Format

A data frame containing 3739 observations on 8 variables:

age numeric. Variable representing age in years (range: 4.8 - 23.6, M = 11.46).

impact numeric. Continuous variable representing severity of and impairment due to mental-health problems at baseline. Higher values indicate higher severity and impairment.

gender factor. Indicator for gender.

emotional factor. Indicator for presence of emotional disorder at baseline.

autism factor. Indicator for presence of autistic disorder at baseline.

conduct factor. Indicator for mental-health service provider.

cluster_id factor. Binarized treatment outcome variable (0 = recovered, 1 = not recovered).

outcome numeric. Variable representing treatment outcome as measured by a total mental-health difficulties score assessed about 6 months after baseline, corrected for the baseline assessment. Higher values indicate poorer outcome.

Details

Dataset was modelled after Edbrooke-Childs et al. (2017), who analyzed a sample of $N = 3,739$ young people who received treatment at one of 13 mental-health service providers in the UK. Note that the data were artificially generated and do not reflect actual empirical findings.

See Also

[lmertree](#), [glmertree](#)

Examples

```
data("MHserviceDemo", package = "glmertree")
summary(MHserviceDemo)
lt <- lmertree(outcome ~ 1 | cluster_id | age + gender + emotional +
              autism + impact + conduct, data = MHserviceDemo)
plot(lt)

gt <- glmertree(factor(outcome > 0) ~ 1 | cluster_id | age + gender +
                emotional + autism + impact + conduct,
                data = MHserviceDemo, family = "binomial")
plot(gt)
```

Index

*Topic **datasets**

DepressionDemo, 2

MHserviceDemo, 6

*Topic **tree**

glmertree, 3

as.party.glmertree (glmertree), 3

as.party.lmertree (glmertree), 3

coef.glmertree (glmertree), 3

coef.lmertree (glmertree), 3

DepressionDemo, 2

glm, 3

glmer, 5

glmerControl, 4

glmertree, 2, 3, 7

glmtree, 5

lmer, 5

lmerControl, 4

lmertree, 2, 7

lmertree (glmertree), 3

lmtree, 5

logLik.glmertree (glmertree), 3

logLik.lmertree (glmertree), 3

MHserviceDemo, 6

mob_control, 4

model.frame.glmertree (glmertree), 3

model.frame.lmertree (glmertree), 3

plot.glmertree (glmertree), 3

plot.lmertree (glmertree), 3

predict.glmertree (glmertree), 3

predict.lmertree (glmertree), 3

print.glmertree (glmertree), 3

print.lmertree (glmertree), 3

ranef.glmertree (glmertree), 3

ranef.lmertree (glmertree), 3

resid.glmertree (glmertree), 3

resid.lmertree (glmertree), 3

residuals.glmertree (glmertree), 3

residuals.lmertree (glmertree), 3

terms.glmertree (glmertree), 3

terms.lmertree (glmertree), 3

VarCorr.glmertree (glmertree), 3

VarCorr.lmertree (glmertree), 3