Package ‘quint’

September 28, 2018

Type Package
Title Qualitative Interaction Trees
Version 2.0.0
Date 2018-09-25
Maintainer Elise Dusseldorp <elise.dusseldorp@fsw.leidenuniv.nl>
Description Grows a qualitative interaction tree. Quint is a tool for subgroup analysis, suitable for data from a two-arm randomized controlled trial.
Depends R (>= 3.0.2), partykit, Formula, rpart, stats
License GPL-2
LazyData TRUE
RoxygenNote 6.1.0
Author Elise Dusseldorp [aut, cre, cph],
Lisa Doove [aut],
Jeanne van de Put [aut],
Cor Ninaber [ctb] (supported with the plot function),
Iven Van Mechelen [aut, cph],
Juan Claramunt Gonzalez [ctb]

NeedsCompilation no
Repository CRAN
Date/Publication 2018-09-28 15:10:02 UTC

R topics documented:

quint-package ......................................................... 2
bcrp ................................................................. 3
plot.quint .......................................................... 4
predict.quint ...................................................... 5
prune.quint ......................................................... 7
quint ................................................................. 8
quint.control ....................................................... 10
quint-package

quint.validate ................................................................. 12
SimData_1 ................................................................. 14
summary.quint ............................................................... 15

Index ................................................................. 17

quint-package R package for Qualitative Treatment-Subgroup Interactions

Description

When two treatment alternatives (say A and B) are available for some problem, one may be inter-
ested in qualitative treatment-subgroup interactions. Such interactions imply the existence of sub-
groups of persons (patients) which are such that in one subgroup Treatment A outperforms Treat-
ment B, whereas the reverse holds in another subgroup. Obviously, this type of interactions is cru-
cial for optimal treatment assignment of future patients. Given baseline characteristics and outcome
data from a two-arm Randomized Controlled Trial (RCT), QAultitative INteraction Trees (QUINT)
is a tool to identify subgroups that are involved in meaningful qualitative treatment-subgroup in-
teractions. The result of QUINT is a tree that partitions the total group of participants (patients)
on the basis of their baseline characteristics into three subgroups (i.e., partition classes): Subgroup
1: Those for whom Treatment A is better than Treatment B (P1), Subgroup 2: Those for whom
Treatment B is better than Treatment A (P2), and Subgroup 3: Those for whom it does not make
any difference (P3).

Details

| Package: quint |
| Type: Package |
| Version: 1.2 |
| Date: 2016-22-04 |
| License: GPL |

This method is suitable for a continuous outcome variable. From version 1.2 onwards the baseline
variables for growing a tree may have numerical or integer values (such as continuous, ordinal or
dichotomous variables) or may be nominal (categorical variables with factors). Previously only
numerical or dichotomous variables were supported. Another new feature of this version is that the
output of a quint object can now also display results for either the raw difference in means or the
effect size with corresponding standard error. This depends on the criterion specified. Furthermore
a predict function predict.quint is newly included in this package. The final new feature is a
validate function quint.validate for estimating the bias (i.e., optimism) of a grown QUINT tree.

From version 2.0 onwards the qualitative treatment-subgroup interaction is checked during the
prune of the tree and not at the begining of QUINT. Furthermore, it is possible to obtain outcomes
from the summary and predict functions when the tree only contains the root node.

The core function of the package is quint.
**bc rp**

**Description**

Data from a three-arm randomized controlled trial. Women with early-stage breast cancer were randomly assigned to a nutrition intervention ($n = 85$), an education intervention ($n = 83$) or standard care ($n = 84$). They were measured before and after treatment. These data contain the baseline measurement and the 9-month follow-up.

**Usage**

`bc rp`

**Format**

A data frame with 252 observations on the following 14 variables:

- `phys t1` physical functioning (from SF-36) at baseline.
- `cesd t1` depression score (CESD) at baseline.
- `phys t3` physical functioning (from SF-36) at 9 months follow-up.
- `cesd t3` depression score (CESD) at 9 months follow-up.
- `negsoct1` negative social interaction at baseline.
- `uncomt1` mitigated communion at baseline.
- `disopt1` dispositional optimism at baseline.
- `comorbid` number of comorbidities (e.g. diabetes, migraines, arthritis, or angina).
age  age at baseline.

wcht1  weight change since diagnosis: yes [1] or no [0].
nationality  Caucasian [1] or not [0].
marital  married [1] or not [0].
tretxt  treatment extensiveness index: lumpectomy without or with one form of adjuvant therapy (radiation or chemo) [-1.77], lumpectomy with radiation and chemotherapy [0.26], mastectomy without or with lumpectomy, and without or with one form of adjuvant therapy [0.56], mastectomy without or with lumpectomy, and radiation and chemotherapy [2.59].

cond  experimental condition: nutrition [1], education [2] or standard care [3].

Details

IMPORTANT: for questions about these data contact Elise Dusseldorp: elise.dusseldorp@fsw.leidenuniv.nl.

Source

The authors thank M.F. Scheier for making his data available.

References


---

**plot.quint**

*Visualisation of a Qualitative Interaction Tree*

**Description**

Plot function for a quint object. The plot shows the result of quint: a binary tree with (a) splitting variable(s) and split point(s). The colors of the leaves of the tree correspond to the final subgroups: Subgroup 1 (P1), those patients for whom the mean treatment outcome (Y) is higher for Treatment A than B, is GREEN; Subgroup 2 (P2), those patients for whom the mean treatment outcome (Y) is higher for Treatment B than A, is RED, and Subgroup 3 (P3), those for whom the mean treatment outcome (Y) is about the same for both treatments, is GREY. Within the leaves the effect size $d$ is displayed, with its 95 percent confidence interval. This effect size is the standardized mean difference between Treatment A and B. The plot function uses the plot method from the package **partykit** of Hothorn & Zeileis (2013).
predict.quint

Usage

## S3 method for class 'quint'
plot(x, digits = 2, ...)

Arguments

x
fitted tree of class quint.
digits
specified number of decimal places of the splitpoints in the graph (default is 2).
... additional arguments to be passed.

Details

For categorical variables we recommend to use short names for levels to avoid overlapping labels at split points.

Author(s)

Cor Ninaber and Elise Dusseldorp

References


See Also

quint.quint.control.bcrp

---

predict.quint Predictions for new data with a QUINT object

Description

Predicts for (new) subjects the treatment subgroups (P1, P2 or P3) based on a fitted quint object. The meaning of the subgroups are based on the two treatment categories used to fit the quint object.

Usage

## S3 method for class 'quint'
predict(object, newdata, type = "class", ...)
## predict.quint

**Arguments**

- **object**: an object of the class “quint”.
- **newdata**: a data frame with data on new subjects for whom predictions should be made. The data frame should contain at least the variables used in the splits of the fitted tree. It is not necessary to include the treatment variable.
- **type**: character string denoting the type of predicted object to be returned. The default is set to type="class": a vector with predicted treatment subgroup classes per subject is returned. If set to “matrix”, a matrix is returned with the leaf and corresponding node of the tree to which a subject is assigned.
- **...**: optional additional arguments.

**Value**

One of the following objects is returned depending on output type specified in the function:

- If type="class": vector of predicted treatment classes for every individual in the data set. Returns NA for subjects with missing values on one or more of the splitting variables.
- If type="matrix": a matrix with predicted locations of subjects within the fitted tree. The leaf numbers are in the first column and the corresponding node numbers in the second column. Returns NA for subjects with missing values on one or more of the splitting variables.

**See Also**

- `quint`, `prune.quint`

**Examples**

```r
data(bcrp)
formula1<- I(cesdt1-cesdt3)~cond | nationality+marital+wcht1+age+
trext+comorbid+disopt1+uncomt1+negsoct1
set.seed(10)
control1<-quint.control(maxl=5,B=2)
quint1<-quint(formula1, data= subset(bcrp,cond<3),control=control1) #Grow a QUINT tree

prquint1<-prune(quint1) #Prune QUINT tree to optimal size

#Predict for the same data set the treatment classes for patients individually:
predquint1<-predict(prquint1, newdata=subset(bcrp,cond<3), type='class')
predquint1
```
**Prune of a Qualitative Interaction Tree**

**Description**

Determines the optimally pruned size of the tree by applying the one standard error rule to the results from the bias-corrected bootstrap procedure.

**Usage**

```r
## S3 method for class 'quint'
prune(tree, pp = 1, ...)
```

**Arguments**

- `tree`: fitted tree of the class `quint`.
- `pp`: pruning parameter, the constant \( c \) to be used in the \( c \times \) standard error rule. The default value is 1.
- `...`: optional additional arguments.

**Details**

The pruning algorithm of `quint` is explained in Dusseldorp & Van Mechelen (2014), Appendix B of the online supplementary material. It is based on the bias-corrected bootstrap pruning procedure (Le Blanc & Crowley, 1993) and the one standard error rule (Breiman, Friedman, Olshen, & Stone, 1984). The one standard error rule for `quint` uses the estimates of the bias-corrected criterion value \( \hat{C} \) and its standard error for each value of \( L \) (= maximum number of leaves). The optimally pruned tree corresponds to the smallest tree with a bias-corrected \( \hat{C} \) higher or equal to the maximum bias-corrected \( \hat{C} \) minus its standard error.

**Value**

Returns an object of class `quint`. The number of leaves of this object is equal to the optimally pruned size of the tree.

**References**


**See Also**

`quint.control`, `quint`
Examples

```r
data(bcrp)
formula2 <- I(cesdt1-cesdt3)-cond | age+trest+uncomt1+ disopt1+negsoct1
#Adjust the control parameters only to save computation time in the example;
#The default control parameters are preferred
control2 <- quint.control(maxl=5,B=2)
set.seed(2) #this enables you to repeat the results of the bootstrap procedure
quint2 <- quint(formula2, data= subset(bcrp,cond<3),control=control2)
quint2pr <- prune(quint2)
summary(quint2pr)
```

---

**quint**

**Qualitative Interaction Trees**

**Description**

This is the core function of the package. It performs a subgroup analysis by QUalitative INteraction Trees (QUINT; Dusseldorp & Van Mechelen, 2014) and is suitable for data from a two-arm randomized controlled trial. Ingredients of the analysis are: one continuous outcome variable \( Y \) (the effect variable), one dichotomous treatment variable \( T \) (indicating two treatment conditions, e.g., A and B), and several background characteristics \( X_1, \ldots, X_J \). These background characteristics are measured at baseline and may have a numeric or ordinal measurement level (i.e., in R a numeric or integer variable) or a nominal measurement level (i.e., in R a factor). They are used to identify the following subgroups (i.e., partition classes): Subgroup 1: Those patients for whom Treatment A is better than Treatment B (P1); Subgroup 2: Those for whom Treatment B is better than Treatment A (P2), and Subgroup 3: Those for whom it does not make any difference (P3).

**Usage**

```r
quint(formula, data, control = NULL)
```

**Arguments**

- `formula`: a description of the model to be fit. The format is \( Y \sim T \mid X_1 + \ldots + X_J \), where the variable before the \( \mid \) represents the dichotomous treatment variable \( T \) and the variables after the \( \mid \) are the baseline characteristics used for partitioning. If the data are in the order \( Y, T, X_1, \ldots, X_J \), no formula is needed. The lay-out of this formula is based on Zeileis & Croissant (2010).
- `data`: a dataframe containing the variables in the model. The treatment variable can be a numeric or a factor variable with two values (or levels).
- `control`: a list with control parameters as returned by `quint.control`. 
Details

The method QUINT uses a sequential partitioning algorithm. The algorithm starts with a tree consisting of a single node, that is, the root node containing all patients. Next, it follows a stepwise binary splitting procedure. This procedure implies that in each step a node, a baseline characteristic, a split of that characteristic, and an assignment of the leaves of the current tree to partition classes 1, 2, and 3 (P1 to P3) are chosen that maximize the partitioning criterion. Note that this means that after each split, all leaves of the tree are re-assigned afresh to the partition classes P1, P2, and P3.

Value

Returns an object of class quint with components:

call  
the call that created the object.

crit  
the partitioning criterion used to grow the tree. The default is the Effect size criterion. Use crit="dm" for the Difference in means criterion.

control  
the control parameters used in the analysis.

fi  
the fit information of the final tree.

si  
the split information of the final tree.

li  
the leaf information of the final tree. Treatment A is denoted with T=1, and treatment B is denoted with T=2. Can display either the output for Difference in Means (crit='dm') or Cohen’s d effect size (crit='es').

data  
the data used to grow the tree.

nind  
an N x L matrix indicating leaf membership.

siboot  
an L x 9 x B array with split information for each bootstrap sample: C_boot = value of C; C_compdif = value of Difference in treatment outcome component; checkdif = indicates if pooled Difference in treatment outcome component in test set (i.e., original sample) is positive, with values: 0 = yes,1 = negative in P1, 2 = negative in P2, 3 = negative in P1 and P2; C_compcard = value of Cardinality component; checkcard = indicates if value of pooled cardinality in test set is zero, with values: 0 = no,1 = zero in P1, 2 = zero in P2, 3 = zero in P1 and P2; opt = value of optimism (C_boot-C_orig).

indexboot  
an N x B matrix indicating bootstrap sample membership.

References


quint.control

Control Parameters for QUINT Algorithm

Description

Various parameters that control aspects of the “quint” algorithm. Appendix A of Dusseldorp & Van Mechelen (2013) gives a detailed overview of the choices that can be made.

Usage

quint.control(crit = "es", maxl = 10, a1 = NULL, a2 = NULL, w = NULL, Bootstrap = TRUE, B = 25, dmin = 0.3)

Arguments

crit the type of difference in treatment outcome used in the partitioning criterion: "es" (Treatment effect sizes) or "dm" (Difference in treatment means).
maxl maximum number of leaves (L) of the tree. Default value is 10.
the minimal sample size of Treatment A ($T = 1$) in a leaf.

The minimal sample size of Treatment B ($T = 2$) in a leaf.

A vector with $w_1$ and $w_2$ representing the weights of, respectively, the Difference in treatment outcome component and the Cardinality component of the partitioning criterion. If crit = "dm", the default value of $w_1$ is $1/log(1 + IQR(Y))$. If crit = "es", the default value of $w_1$ is $1/log(1 + 3)$. The default of $w_2$ is $1/log(0.50N)$.

**Bootstrap**

whether the bias-corrected bootstrap procedure should be performed. The default is TRUE.

**B**

the number of bootstrap samples to be drawn. The default is 25. We recommend a number of bootstraps of at least 25.

**dmin**

the minimum absolute standardized mean difference in treatment outcome in each of the two leaves after the first split of the tree. This value is used to check whether a qualitative interaction is present in the data (the qualitative interaction condition); dmin controls the balance between Type I error and Type II error. The default value of dmin is 0.30.

**Value**

A list containing the options.

**References**


**See Also**

quint

**Examples**

data(bcrp)
formula1<- I(cesdt1-cesdt3)~cond | nationality+marital+wcht1+age+trest+comorbid+disopt1+uncomt1+negsoct1
#Specify the Difference in treatment outcome as Difference in means
#and skip the bias-corrected bootstrap procedure
#and change the maximum number of leaves
control3<-quint.control(crit="dm",Bootstrap=FALSE,maxl=3)
quint3<-quint(formula1, data= subset(bcrp,cond<3),control=control3)
summary(quint3)

#Set number of bootstrap samples at 30
control4<-quint.control(B=30)

#Set minimal sample size in each treatment group at 5
Validation of a Qualitative Interaction Tree

Description

A bootstrap-based validation procedure to estimate the optimism in the effect sizes of a QUINT tree which gives insight in the generalizability of the results.

Usage

quint.validate(object, B = 10, allresults = FALSE)

Arguments

object a (pruned) QUINT tree object of class quint.
B number of bootstrap samples. Default number is 10; for better accuracy B=1000 is recommended.
allresults option to return an extended list of output. Default is set to FALSE. See Value section for details.

Details

In this procedure bootstrap trees are grown of the same leaf size as the (pruned) QUINT tree. The bootstrap samples are drawn from the data used to grow the original tree. For every bootstrap tree the largest and smallest (i.e., largest negative) treatment mean differences (or treatment effect sizes) of two leaves are saved. Treatment mean differences in the leaves are then predicted using the original data set as input for each bootstrapped tree. From these predictions, the largest and smallest treatment mean differences are saved. For each bootstrap tree, the largest predicted treatment effect is subtracted from the largest treatment effect in the bootstrap sample. The average of these values is the bias (i.e., the optimism) for the largest treatment effects. This is done likewise for the smallest treatment effects. Subsequently, the bias is computed as the difference between the bias for the largest effects minus the bias for the smallest effects.

The details of this validation procedure are described in Appendix C of Dusseldorp & Van Mechelen (2014).

Value

Returns a list with the following components:

estopt the estimated optimism for either the treatment effect size (biasd) or the raw treatment mean difference (biasdif).
a data frame with leaf information output similar to the leaf information output of the (pruned) QUINT tree object. An extra column is added for the bias-corrected differences in treatment outcomes (d or diff). The bias-corrected values are only computed for the leaves with the most extreme values, i.e. the largest and smallest treatment effects. Hence, the other leaves get the value NA in this column.

optd  
a matrix with computed estimated optimism of the treatment effect size per bootstrapped tree. The first column contains the difference between the largest and smallest effect size of the bootstrapped tree. The second column contains the difference between the largest and smallest predicted effect size. Returned when allresults is set to TRUE and crit='es' is specified in the QUINT object.

optdif  
a matrix with computed estimated optimism of the raw mean difference bootstrapped tree. The first column contains the difference between the largest and smallest raw mean difference of the bootstrapped tree. The second column contains the difference between the largest and smallest predicted raw mean difference. Returned when allresults is set to TRUE and crit='es' is specified in the QUINT object.

resultd  
a vector with the estimated overall mean optimism, the mean bias for the smallest and for the largest effect size. Returned when allresults is set to TRUE and crit="es".

resultdif  
a vector with the estimated overall mean optimism, the mean bias for the smallest and largest raw mean difference. Returned when allresults is set to TRUE and crit="dm".

References


See Also

*quint, prune.quint, quint.control*

Examples

data(bcrp)
formula<- I(cesdt1-cesdt3)~cond | nationality+marital+wcht1+age+text+comorbid+disopt1+uncomt1+negsoct1

set.seed(10)
control1<-quint.control(maxl=4,B=2)
quint1<-quint(formula1, data= subset(bcrp,cond<3),control=control1) #Grow a QUINT tree

prquint1<-prune(quint1) #Prune tree to optimal size

set.seed(3)
valquint1<-quint.validate(prquint1, B = 5) #estimate the optimism by bootstrapping 5 times
valquint1
Description

Data generated sampling from a multivariate normal distribution using as parameters \( samp = 150 \), \( es = 1.0, J = 20, \rho = 0 \) where \( samp \) is the sample size, \( es \) is the effect size, \( J \) is the number of covariates, and \( \rho \) is the correlation between pairs of covariates. The treatment condition (A) is sampled from a binomial distribution with \( p = 0.5 \) and the treatment outcome (Y) is obtained using the following regression \( Y = 1 + (0.25X_1) + (0.25X_2) - (0.25X_5) - (esx((A - 1)^2)) + \text{error} \) where the error is sampled from a standard normal distribution.

Usage

SimData_1

Format

A data frame with 150 observations on the following variables:

- Y: Outcome variable. It is obtained by regression according to the previously explained formula.
- A: This is the treatment condition variable. It is sampled from a binomial distribution.
- X1: Simulated covariate. It is sampled from a multivariate normal distribution.
- X2: Simulated covariate. It is sampled from a multivariate normal distribution.
- X3: Simulated covariate. It is sampled from a multivariate normal distribution.
- X4: Simulated covariate. It is sampled from a multivariate normal distribution.
- X5: Simulated covariate. It is sampled from a multivariate normal distribution.
- X6: Simulated covariate. It is sampled from a multivariate normal distribution.
- X7: Simulated covariate. It is sampled from a multivariate normal distribution.
- X8: Simulated covariate. It is sampled from a multivariate normal distribution.
- X9: Simulated covariate. It is sampled from a multivariate normal distribution.
- X10: Simulated covariate. It is sampled from a multivariate normal distribution.
- X11: Simulated covariate. It is sampled from a multivariate normal distribution.
- X12: Simulated covariate. It is sampled from a multivariate normal distribution.
- X13: Simulated covariate. It is sampled from a multivariate normal distribution.
- X14: Simulated covariate. It is sampled from a multivariate normal distribution.
- X15: Simulated covariate. It is sampled from a multivariate normal distribution.
- X16: Simulated covariate. It is sampled from a multivariate normal distribution.
- X17: Simulated covariate. It is sampled from a multivariate normal distribution.
- X18: Simulated covariate. It is sampled from a multivariate normal distribution.
- X19: Simulated covariate. It is sampled from a multivariate normal distribution.
- X20: Simulated covariate. It is sampled from a multivariate normal distribution.
- gopt: Variable indicating whether the data is simulated such that there is qualitative interaction (0) or there is no qualitative interaction (1)
### summary.quint

**Summarizing Qualitative Interaction Tree Information**

#### Description

Summary method for an object of class `quint`.

#### Usage

```r
## S3 method for class 'quint'
summary(object, digits = 2, ...)
```

#### Arguments

- `object`: a `quint` object. This can be the output of `quint`.
- `digits`: specified number of decimal places (default is 2).
- `...`: optional additional arguments.

#### Details

This function is a method for the generic function `summary` for class `quint`. It extracts the following essential components from a `quint` object: 1) Specification of the partitioning criterion; 2) Fit information; 3) Split information, and 4) Leaf information.

#### Value

prints a summarized version of the `quint` output.

#### Examples

```r
data(bcrp)
formula <- I(cesdt1~cesdt3)~cond | nationality+marital+wcht1+
age+text+comorbid+disopt1+uncomt1+negsoct1
control <- quint.control(maxl=5, Bootstrap=FALSE)
quint1 <- quint(formula, data=subset(bcrp, cond<3), control=control1)
summary(quint1)

# Example with only root node tree as outcome
data(SimData_1)
formula <- Y~A | X1+X2+X3+X4+X5
# Adjust the control parameters only to save computation time in the example;
# The default control parameters are preferred
control <- quint.control(maxl=5, B=2)
set.seed(2) # this enables you to repeat the results of the bootstrap procedure
```
quint_1 <- quint(formula, data = SimData_1, control = control)
quint_1pr <- prune(quint_1)
summary(quint_1pr)
Index

*Topic `as.party`
  plot.quint, 4
*Topic `cluster`
  quint, 8
*Topic `datasets`
  bcrp, 3
  SimData_1, 14
*Topic `package`
  quint-package, 2
*Topic `plot`
  plot.quint, 4
*Topic `summary`
  summary.quint, 15
*Topic `tree`
  prune.quint, 7
  quint, 8

bcrp, 3, 5, 10
plot.quint, 4
predict.quint, 3, 5
prune.quint, 3, 6, 7, 10, 13

quint, 2, 3, 5–7, 8, 11, 13, 15
quint-package, 2
quint.control, 3, 5, 7, 8, 10, 10, 13
quint.validate, 3, 12

SimData_1, 14
summary.quint, 3, 10, 15