Tree-based Multilevel Models

Background

- 2014: IPDMA dataset
  - multiple studies comparing CBT and PHA, numerous potential predictors and moderators of treatment effect
- Model-based recursive partitioning (Zeileis, Hothorn & Hornik, 2008)
- How to deal with multilevel structure?

Contents

1) Introduction to trees
   - Trees vs. linear models
   - R ex. #1: Predicting depression using personality scales

2) Trees for multilevel and longitudinal data (mixed-effects trees)
   - R ex. #2: Predicting treatment outcomes in a multi-center dataset
   - R ex. #3: Detecting treatment-subgroup interactions in a multi-center dataset

Assumed skills and knowledge

- Some knowledge about and skill in fitting and interpreting generalized linear mixed-effects models (GLMs and GLMMs)
- Some experience with fitting these models in R
- Slides, datasets and commented code at https://github.com/marjoleinF/TSIL2018
- Copy and execute code yourself, from file 'Fitting_mixed-effects_trees_in_R.pdf'
  - Feel free to tweak, tune, adjust and comment

Advantages of tree methods

1. Data-analytic flexibility
   - Predictor variables may be categorical, ordinal or continuous
   - No (or very few) assumptions about data distribution
     - E.g., no assumptions of linearity or additivity; interactions and non-linear effects automatically accomodated
   - Can deal with many potential predictors (p > N no problem)

2. Easy to interpret and apply
   - Direct identification of subgroups
   - Easy to use in decision making (no formulas, no computation)
   - Not all variables in model need be assessed for making decision

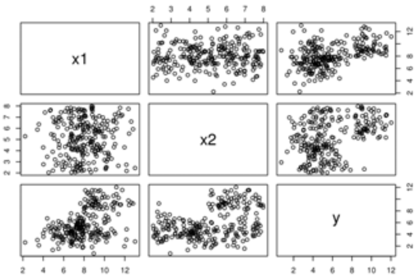
Disadvantage of tree methods

3. Instability: Small changes in data may yield very different tree

   But:
   - Traditional (generalized) linear models are also unstable, changing tree structure just looks much worse than changing coefficients in linear model
   - Just as with (generalized) linear models, mixed-effects trees are more stable and accurate
   - Tree ensembles provide state-of-the-art predictive accuracy, but also suboptimal interpretability
     - Prediction rule ensembles strike balance between accuracy and interpretability. See R package pre (Fokkema & Christoffersen, 2018).

Traditional (generalized) linear model vs. trees

```
toy_data <- read.table("toy data.txt")
plot(toy_data)
```

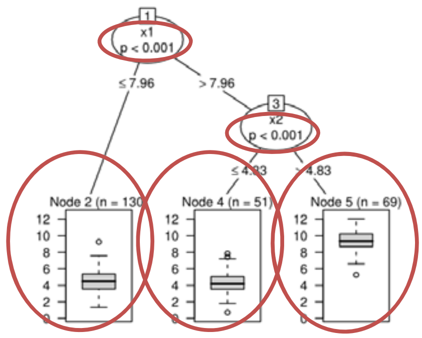


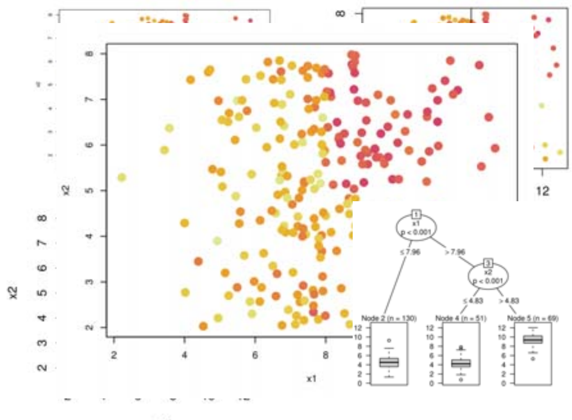

```
linear_model <- lm(y ~ x1 + x2, data = toy_data)
summary(linear_model)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = toy_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7997 -1.5850 -0.0328  1.4018  5.2515
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.00288    0.69683  -2.874   0.0044 **
## x1           0.61208    0.06961   8.793 2.54e-16 ***
## x2           0.56791    0.08005   7.095 1.37e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

$$\hat{y} = -2 + 0.61\ x1 + 0.57\ x2$$

```
library("partykit")
lm_tree <- lmtree(y ~ 1 | x1 + x2, data = toy_data)
plot(lm_tree, gp = gpar(cex = .75))
```





Example 1: Predicting depression with personality scales

```
carrillo <- read.table("carrillo data.txt")
names(carrillo)
```

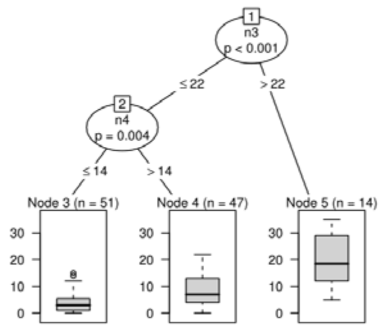
```
##  [1] "n1"      "n2"      "n3"      "n4"      "n5"      "n6"      "ntot"
##  [8] "e1"      "e2"      "e3"      "e4"      "e5"      "e6"      "etot"
## [15] "open1"   "open2"   "open3"   "open4"   "open6"   "opentot" "altot"
## [22] "contot"  "bdi"     "sexo"    "edad"    "open5"
```

```
dim(carrillo)
```

```
## [1] 112  26
```

```
car_tree <- lmtree(bdi ~ 1 | ., data = carrillo)
```

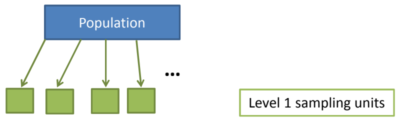
```
plot(car_tree, gp = gpar(cex = .75))
```



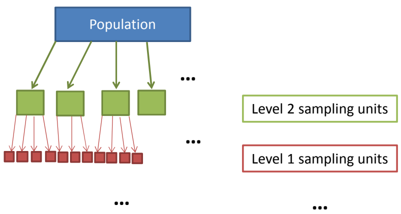
```
car_tree

## Linear model tree
##
## Model formula:
## bdi ~ 1 | .
##
## Fitted party:
## [1] root
## |   [2] n3 <= 22
## |   |   [3] n4 <= 14: n = 51
## |   |       (Intercept)
## |   |          3.72549
## |   |   [4] n4 > 14: n = 47
## |   |       (Intercept)
## |   |          8.808511
## |   [5] n3 > 22: n = 14
## |       (Intercept)
## |          20.64286
##
## Number of inner nodes:    2
## Number of terminal nodes: 3
## Number of parameters per node: 1
## Objective function (residual sum of squares): 3622.648
```

'Fixed-effects' sampling structure



Multilevel or longitudinal data



Multilevel or longitudinal data

Usual mixed-effects model: $\boldsymbol{y = X\beta + Zb + \varepsilon}$

- Fixed-effects (population level) part $\boldsymbol{X\beta}$
- Random-effects (cluster-specific) part $\boldsymbol{Zb}$

Mixed-effects model trees (Fokkema et al., in press):

- Fixed-effects part $\boldsymbol{X\beta}$ is replaced by a tree
- Random effects $\boldsymbol{b}$ estimated as usual (globally)

```
library("glmertree")

UKMH_data <- read.table("UK_MH_mimic data.txt")
dim(UKMH_data)

## [1] 3739    8

head(UKMH_data)

##    age impact gender emotional autism conduct cluster_id outcome
## 1 16.0    4.9 female       yes     no      no          8    -0.2
## 2  9.4    4.4 female       yes     no      no          1     0.2
## 3 12.6    2.5   male        no     no      no          2    -0.6
## 4 13.5    3.7   male       yes     no     yes         13     0.0
## 5 12.7    0.9 female       yes     no      no         12     0.7
## 6 11.0    3.6   male        no     no      no          7     0.0

UKMH_tree <- lmertree(outcome ~ 1 | cluster_id | age + gender + emotional +
                      autism + impact + conduct, data = UKMH_data)
```

Specifying more complex random-effects structures

E.g., nested cluster indicators:

```
outcome ~ 1 | ( (1 | service / department) ) | age + gender +
  emotional + autism + impact + conduct
```

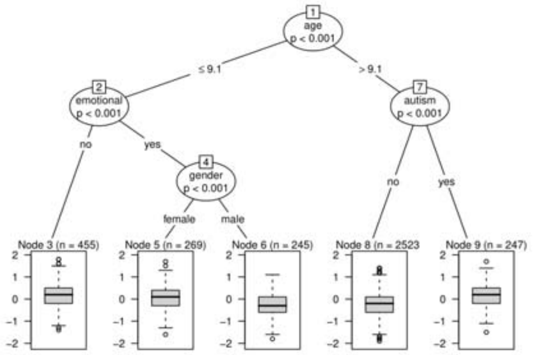
E.g., crossed cluster indicators:

```
outcome ~ 1 | ( (1 | cluster_id1) + (1 | cluster_id2) ) | age + gender +
  emotional + autism + impact + conduct
```

E.g., random slopes (in addition to random intercepts):

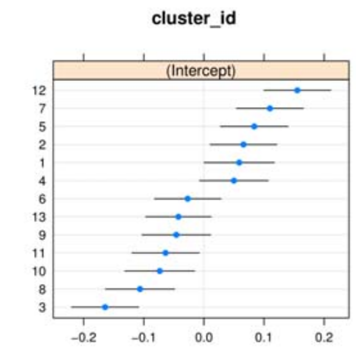
```
outcome ~ 1 | (1 + baseline | cluster_id) | age + gender +
  emotional + autism + impact + conduct
```

```r
plot(UKMH_tree, which = "tree", gp = gpar(cex = .75))
```




```r
plot(UKMH_tree, which = "ranef", width = 0.5)
```

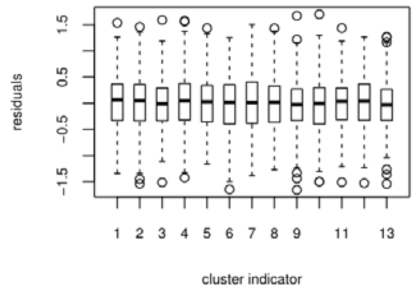
```
## $cluster_id
```



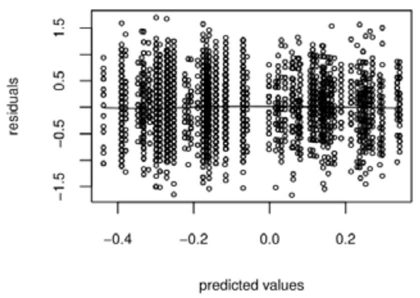
Checking (mis-)specification

- Estimation of random effects introduces assumptions about the data distribution and thereby possibilities for misspecification
- Misspecification may be difficult to detect, but plotting and inspecting the data can be of great help:
  - Plotting the tree and random effects already provides a check on (mis-)specification of the model
    - E.g., node-specific means and variances, outliers?
  - Residuals should also be checked
    - E.g., for heteroscedasticity

```r
resids <- residuals(UKMH_tree)
plot(x = factor(UKMH_data$cluster_id), y = resids,
     xlab = "cluster indicator", ylab = "residuals",
     boxwex = .5, cex.lab=.75, cex.axis = .75)
```




```r
preds <- predict(UKMH_tree)
scatter.smooth(x = preds, y = resids, xlab = "predicted values",
               ylab = "residuals", cex = .5, cex.lab = .75, cex.axis = .75)
```



Detecting treatment outcome moderators and predictors in multilevel data

```r
SMART_data <- read.table("SMART_mimic data.txt")
dim(SMART_data)
```

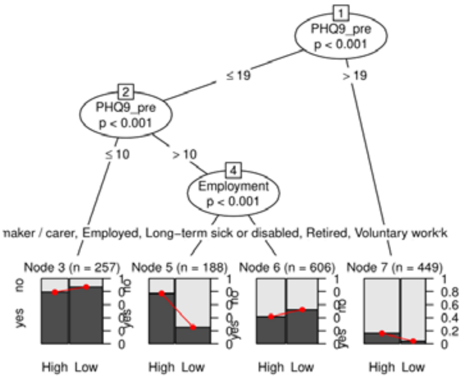
```
## [1] 1500   13
```

```r
names(SMART_data)
```

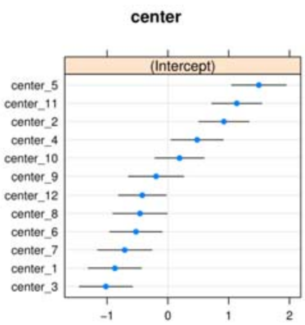
```
##  [1] "recovered"  "Treatment"  "Age"        "Gender"     "Ethnicity"
##  [6] "Diagnosis"  "Employment" "Disability" "PHQ9_pre"   "GAD7_pre"
## [11] "WSAS_pre"   "Medication" "center"
```

```r
SMART_tree <- glmertree(recovered ~ Treatment | center | Age + PHQ9_pre +
                          GAD7_pre + WSAS_pre + Gender + Ethnicity +
                          Diagnosis + Employment + Disability + Medication,
                        data = SMART_data, family = binomial)
```

```r
plot(SMART_tree, which = "tree", gp = gpar(cex = .75))
```




```
## Fitted party:
## [1] root
## |   [2] PHQ9_pre <= 19
## |   |   [3] PHQ9_pre <= 10: n = 257
## |   |         (Intercept) TreatmentLow
## |   |           1.5288250    0.5570944
## |   |   [4] PHQ9_pre > 10
## |   |   |   [5] Employment in Homemaker / carer, Student, Unemployed job seeker, Unemployed, not see
## |   |   |         (Intercept) TreatmentLow
## |   |   |            1.208380    -2.467001
## |   |   |   [6] Employment in Employed, Long-term sick or disabled, Retired, Voluntary work: n = 606
## |   |   |         (Intercept) TreatmentLow
## |   |   |          -0.3028141    0.4358695
## |   [7] PHQ9_pre > 19: n = 449
## |         (Intercept) TreatmentLow
## |           -1.873063    -1.530169
```



Concluding remarks

- Package glmertree offers quite some flexibility, you can adjust the model formula in many ways to suit your data and research question
- Should covariate be in the node-specific model or a partitioning variable?
  - Node-specific model is the confirmatory part
    - E.g., in RCT data, treatment should definitely be included as a predictor in the node-specific model
  - Partitioning variables are the exploratory part
    - E.g., possible predictors or moderators of response variable
- Current and future developments:
  - Methods for cluster-level partitioning variables
    - By default, partitioning variables assumed to be on lowest unit-level
    - Yields too high Type II error for cluster-level variables, for which ‘cluster’ argument should be employed (implemented in development version on R-Forge)
  - Partitioning growth curve models
  - Requests / questions for new stuff: let me know!

Dataset papers:

Carrillo, J. M., Rojo, N., Sanchez-Bernardos, M. L., Avia, M. D. (2001). Openness to experience and depression. *European Journal of Psychological Assessment, 17*(2), 130-136.

Edbrooke-Childs, J., Macdougall, A., Hayes, D., Jacob, J., Wolpert, M., & Deighton, J. 9 (2017). Service-level variation, patient-level factors, and treatment outcome in those seen by child mental health services. *European Child & Adolescent Psychiatry, 26(6*), 715-722.

Methodological papers:

Fokkema, M., Smits, N., Zeileis, A., Hothorn, T. & Kelderman, H. (in press). Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. *Behavior Research Methods*. url: https://link.springer.com/article/10.3758/s13428-017-0971-x

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics, 17* (2), 492-514.

R packages:

Fokkema, M. & Christoffersen (2017). pre: Prediction Rule Ensembles. R package version 0.5. Stable: https://cran.r-project.org/package=glmertree. Development: https://github.com/marjoleinF/pre

Hothorn, T., Seibold, H. & Zeileis, A. (2018). partykit: A Toolkit for Recursive Partytioning. R package version 1.2-2. url: https://cran.r-project.org/package=partykit

Zeileis, A. & Fokkema, M. (2017). glmertree: Generalized Linear Mixed Model Trees. R package version 0.1. Stable: https://cran.r-project.org/package=glmertree. Development: https://r-forge.r-project.org/R/?group_id=261