partykit: A Toolbox for Recursive Partytioning

Torsten Hothorn[1] **and Achim Zeileis**[2]

1. Institut für Statistik, Ludwig-Maximilians-Universität München, Germany
   `Torsten.Hothorn@R-project.org`
2. Department of Statistics and Mathematics, Wirtschaftsuniversität Wien, Austria
   `Achim.Zeileis@R-project.org`

Keywords: Regression Trees, Visualization, CTree, CHAID, Weka, PMML.

Recursive partitioning methods, or simply "trees", are simple yet powerful methods for capturing regression relationships. Since the publication of the automated interaction detection (AID) algorithm in 1964, many extensions, modifications, and new approaches have been suggested in both the statistics and machine learning communities. Most of the standard algorithms are available to the R user, e.g., through packages **rpart** (Therneau *et al.*, 2009), **RWeka** (Hornik *et al.*, 2009), **party** (Hothorn *et al.*, 2009), or **mvpart** (De'ath, 2007).

However, no common infrastructure is available for representing trees fitted by different packages. Consequently, the capabilities for extraction of information—such as predictions, printed summaries, or visualizations—vary between packages and come with somewhat different user interfaces. Furthermore, extensions or modifications often require considerable programming effort, e.g., if the median instead of the mean of a numerical response should be predicted in each leaf of an '`rpart`' tree. Similarly, implementations of new tree algorithms might also require new infrastructure if they have features not available in the above-mentioned packages, e.g., multi-way splits or more complex models in the leafs.

To overcome these difficulties, the **partykit** package (Hothorn and Zeileis, 2009) offers a unified representation of tree objects along with `predict()`, `print()`, and `plot()` methods. Trees are represented through a new flexible class '`party`' which can, in principle, capture all trees mentioned above but can also accommodate multi-way or functional splits, as well as complex models in (leaf) nodes. The package is currently under development at R-Forge but already provides conversion methods for trees of classes '`rpart`', '`J48`', and '`pmmlTreeModel`' as well as a re-implementation of conditional inference trees (Hothorn *et al.*, 2006).

In our presentation, we will only sketch details of these classes and corresponding methods and focus on applications of the toolkit including extended visualizations for '`rpart`' or '`J48`' objects, fast predictions on millions of new observations, and a new implementation of the classical CHAID algorithm.

References

De'ath G (2007). ***mvpart**: Multivariate Partitioning*. R package version 1.2-6, URL `http://CRAN.R-project.org/package=mvpart`.

Hornik K, Zeileis A, Hothorn T, Buchta C (2009). ***RWeka**: An R Interface to **Weka***. R package version 0.3-16, URL `http://CRAN.R-project.org/package=RWeka`.

Hothorn T, Hornik K, Strobl C, Zeileis A (2009). ***party**: A Laboratory for Recursive Partytioning*. R package version 0.9-995, URL `http://CRAN.R-project.org/package=party`.

Hothorn T, Hornik K, Zeileis A (2006). "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics*, **15**(3), 651–674. doi:10.1198/106186006X133933.

Hothorn T, Zeileis A (2009). ***partykit**: A Toolkit for Recursive Partytioning*. R package version 0.0-1, URL `http://R-Forge.R-project.org/projects/partykit/`.

Therneau TM, Atkinson B, Ripley BD (2009). ***rpart**: Recursive Partitioning*. R package version 3.1-42, URL `http://CRAN.R-project.org/package=rpart`.