Model-Based Recursive Partitioning for Detecting Interaction Effects in Subgroups

Achim Zeileis, Torsten Hothorn, Kurt Hornik

http://eeecon.uibk.ac.at/~zeileis/
Overview

- Motivation: Trees, leaves, and branches
- Model-based recursive partitioning
  - Model estimation
  - Tests for parameter instability
  - Segmentation
  - Pruning
- Application: Treatment effect for chronic disease
- Summary
Motivation: Trees


- **Data models**: Stochastic models, typically parametric.
- **Algorithmic models**: Flexible models, data-generating process unknown.

**Example**: Recursive partitioning models dependent variable $Y$ by “learning” a partition w.r.t explanatory variables $Z_1, \ldots, Z_l$.

**Key features**:

- Predictive power in nonlinear regression relationships with “automatic interaction detection”.
- Interpretability (enhanced by visualization), i.e., no “black box” methods.
Motivation: Leaves

Typically: Simple models for univariate $Y$, e.g., mean or proportion.

Examples: CART and C4.5 in statistical and machine learning, respectively.

Problems: For classical tree algorithms.
- No concept of “significance”, possibly biased variable selection.
- No complex (parametric) models in leaves.
- Many different tree algorithms for different types of data.

Here: Synthesis of parametric data models and algorithmic tree models.
- Fitting local models by partitioning of the sample space.
- Based on statistical hypothesis tests for parameter instability.
Motivation: Branches

**Base algorithm:** Growth of branches from the roots to the leaves of the tree follows a generic *recursive partitioning* algorithm.

1. Fit a (possibly very simple) model for the response $Y$.
2. Assess association of $Y$ and each $Z_j$.
3. Split sample along the $Z_j^*$ with strongest association: Choose breakpoint with highest improvement of the model fit.
4. Repeat steps 1–3 recursively in the subsamples until some stopping criterion is met.

**Here:** Segmentation (3) of parametric models (1) with additive objective function using parameter instability tests (2) and associated statistical significance (4).
Model-based recursive partitioning: Estimation

**Models:** \( \mathcal{M}(Y, \theta) \) with (potentially) multivariate observations \( Y \in \mathcal{Y} \) and \( k \)-dimensional parameter vector \( \theta \in \Theta \).

**Parameter estimation:** \( \hat{\theta} \) by optimization of objective function \( \psi(Y, \theta) \) for \( n \) observations \( Y_i (i = 1, \ldots, n) \):

\[
\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^{n} \psi(Y_i, \theta).
\]

**Special cases:** Maximum likelihood (ML), weighted and ordinary least squares (OLS and WLS), quasi-ML, and other M-estimators.

**Central limit theorem:** If there is a true parameter \( \theta_0 \) and given certain weak regularity conditions, \( \hat{\theta} \) is asymptotically normal with mean \( \theta_0 \) and sandwich-type covariance.
Model-based recursive partitioning: Estimation

**Estimating function:** $\hat{\theta}$ can also be defined in terms of

$$\sum_{i=1}^{n} \psi(Y_i, \hat{\theta}) = 0,$$

where $\psi(Y, \theta) = \partial \Psi(Y, \theta)/\partial \theta$.

**Idea:** In many situations, a single global model $M(Y, \theta)$ that fits all $n$ observations cannot be found. But it might be possible to find a partition w.r.t. the variables $Z = (Z_1, \ldots, Z_l)$ so that a well-fitting model can be found locally in each cell of the partition.

**Tool:** Assess parameter instability w.r.t. to partitioning variables $Z_j \in \mathcal{Z}_j$ ($j = 1, \ldots, l$).
Model-based recursive partitioning: Tests

Generalized M-fluctuation tests capture instabilities in $\hat{\theta}$ for an ordering w.r.t $Z_j$.

**Basis:** Empirical fluctuation process of cumulative deviations w.r.t. to an ordering $\sigma(Z_{ij})$.

$$ W_j(t, \hat{\theta}) = \hat{V}^{-1/2} n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \psi(Y_{\sigma(Z_{ij})}, \hat{\theta}) \quad (0 \leq t \leq 1) $$

**Functional central limit theorem:** Under parameter stability $W_j(\cdot, \hat{\theta}) \xrightarrow{d} W^0(\cdot)$, where $W^0$ is a $k$-dimensional Brownian bridge.
Model-based recursive partitioning: Tests

**Test statistics:** Scalar functional $\lambda(W_j)$ that captures deviations from zero.

**Null distribution:** Asymptotic distribution of $\lambda(W^0)$.

**Special cases:** Class of test encompasses many well-known tests for different classes of models. Certain functionals $\lambda$ are particularly intuitive for numeric and categorical $Z_j$, respectively.

**Advantage:** Model $M(Y, \hat{\theta})$ just has to be estimated once. Empirical estimating functions $\psi(Y_i, \hat{\theta})$ just have to be re-ordered and aggregated for each $Z_j$. 
Splitting numeric variables: Assess instability using supLM statistics.

\[
\lambda_{\text{supLM}}(W_j) = \max_{i=\bar{i}, \ldots, \bar{i}} \left( \frac{i}{n} \cdot \frac{n-i}{n} \right)^{-1} \left\| W_j \left( \frac{i}{n} \right) \right\|_2^2.
\]

Interpretation: Maximization of single shift LM statistics for all conceivable breakpoints in [i, \bar{i}].

Limiting distribution: Supremum of a squared, k-dimensional tied-down Bessel process.
Model-based recursive partitioning: Tests

Splitting categorical variables: Assess instability using $\chi^2$ statistics.

$$\lambda_{\chi^2}(W_j) = \sum_{c=1}^{C} \frac{n}{|I_c|} \left\| \Delta_{I_c} W_j \left( \frac{i}{n} \right) \right\|_2^2$$

Feature: Invariant for re-ordering of the $C$ categories and the observations within each category.

Interpretation: Captures instability for split-up into $C$ categories.

Limiting distribution: $\chi^2$ with $k \cdot (C - 1)$ degrees of freedom.
Model-based recursive partitioning: Segmentation

**Goal:** Split model into $b = 1, \ldots, B$ segments along the partitioning variable $Z_j$ associated with the highest parameter instability. Local optimization of

$$\sum_b \sum_{i \in I_b} \psi(Y_i, \theta_b).$$

$B = 2$: Exhaustive search of order $O(n)$.

$B > 2$: Exhaustive search is of order $O(n^{B-1})$, but can be replaced by dynamic programming of order $O(n^2)$. Different methods (e.g., information criteria) can choose $B$ adaptively.

**Here:** Binary partitioning.
Pruning: Avoid overfitting.

Pre-pruning: Internal stopping criterion. Stop splitting when there is no significant parameter instability.

Post-pruning: Grow large tree and prune splits that do not improve the model fit (e.g., via cross-validation or information criteria).

Here: Pre-pruning based on Bonferroni-corrected $p$ values of the fluctuation tests.
Application: Treatment effect for chronic disease

Task: Identify groups of chronic disease patients with different treatment effects.

Source: Anonymized data from consulting project.

Model: Logistic regression estimated by maximum likelihood.
- Response: Improvement (yes/no) of chronic disease for 1354 patients after treatment over several weeks.
- Regressor: Treatment (active drug/placebo).
- Partitioning variables: 11 variables that describe disease status of patients. Lower values indicate more severe forms of the disease.

Result: Treatment most effective for certain intermediate forms.
Application: Treatment effect for chronic disease

Node 2 (n = 689)

- **Placebo**
  - Yes
  - No

- **Drug**
  - Yes
  - No

Node 4 (n = 527)

- **Placebo**
  - Yes
  - No

- **Drug**
  - Yes
  - No

Node 5 (n = 138)

- **Placebo**
  - Yes
  - No

- **Drug**
  - Yes
  - No

**Risk 7**

- \( p = 0.004 \)
- \( \leq 0 \rightarrow \text{Yes} \)
- \( > 0 \rightarrow \text{No} \)

**Risk 5**

- \( p = 0.042 \)
- \( \leq 3.765 \rightarrow \text{Yes} \)
- \( > 3.765 \rightarrow \text{No} \)
Application: Treatment effect for chronic disease

Model-based recursive partitioning:
- Coefficient estimates for regressors.
- Parameter instability tests for partitioning variables (bold = significant at adjusted 5% level, underlined = smallest $p$ value).

<table>
<thead>
<tr>
<th>Regressors</th>
<th>Partitioning variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>(const.)</td>
<td>treatment</td>
</tr>
<tr>
<td>1</td>
<td>−0.99</td>
</tr>
<tr>
<td>2</td>
<td>−1.15</td>
</tr>
<tr>
<td>3</td>
<td>−0.82</td>
</tr>
<tr>
<td>4</td>
<td>−0.72</td>
</tr>
<tr>
<td>5</td>
<td>−1.13</td>
</tr>
</tbody>
</table>
Summary

- Synthesis of parametric data models and algorithmic tree models.
- Based on modern class of parameter instability tests.
- Aims to minimize clearly defined objective function by greedy forward search.
- Can be applied to general class of parametric models: generalized linear models, psychometric models (e.g., Rasch, Bradley-Terry), models for location and scale, etc.
- Automatic interaction detection of effects in subgroups.
- Alternative to traditional means of model specification, especially for variables with unknown association.
- Object-oriented implementation freely available: Extension for new models requires some coding, though.
