Prediction rule ensembles for multilevel data

Marjolein Fokkema, Leiden University

Prediction rule ensembles (PREs)

Goal: Predict y with sparse set of rules

Example: *if* [color == blue & waist circumference > 50] *then* [log odds +6.3]

Tasks: 1) Find rules; 2) Estimate coefficients

RuleFit algorithm (Friedman & Popescu, 2008)

1) Construct tree ensemble:
   - Boosting with subsampling
2) Construct initial ensemble:
   - Include each node from each tree as a rule, and/or
   - Include predictor variables
3) Select final ensemble by sparse regression in training data
   - Using lasso, ridge or elastic net penalty

Final ensemble:

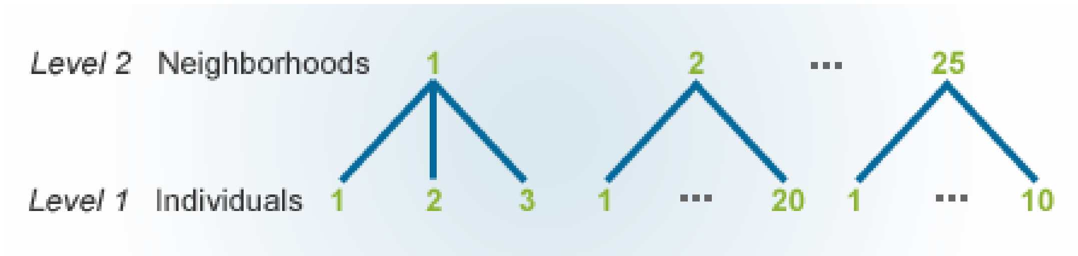
Rule, e.g.:
$f_m(\mathbf{x}) = I(x_1 > 5) \cdot I(x_4 \leq 10)$

$$F(\mathbf{x}) = \hat{a}_0 + \sum_{m=1}^{M} \hat{a}_m f_m(\mathbf{x})$$

Linear term, e.g.:
$f_m(\mathbf{x}) = x_2$

Multilevel data

Multilevel structure:



Possible approaches:

-> Ignore clustered structure

-> Sample level-2 instead of level-1 units

-> Estimate random effects

Mixed-effects PRE

'Fixed effects' PRE:

$$F(\mathbf{x}) = a_0 + \sum_{m=1}^{M} a_m f_m(\mathbf{x})$$

Add random intercepts:

$$F(\mathbf{x}, \mathbf{z}) = a_0 + \sum_{m=1}^{M} a_m f_m(\mathbf{x}) + \mathbf{z}^{\mathrm{T}}\mathbf{b}$$

Task: Estimate rules, $\mathbf{a}$ and $\mathbf{b}$ values

Estimation

$$F(\mathbf{x}, \mathbf{z}) = a_0 + \sum_{m=1}^{M} a_m f_m(\mathbf{x}) + \mathbf{z}^{\mathrm{T}}\mathbf{b}$$

Iterative approach:

step 0) Initialize with setting $\hat{\mathbf{b}} = \mathbf{0}$

step 1) Estimate $f_m(\mathbf{x})$ and $\hat{\mathbf{a}}$ given current $\hat{\mathbf{b}}$

step 2) Estimate $\hat{\mathbf{b}}$ given current $\hat{\mathbf{a}}$

step 3) Repeat steps 2 & 3 until convergence

Works well for single trees and random forests (e.g., Fokkema et al., in press; Hajjem et al., 2014)

PRE methods for clustered data

pre: ignores multilevel structure

pre_cs: cluster-level sampling

pre_m_part: rules estimated as in pre + random effects included in estimation of final ensemble

pre_m_full: estimate rules with mixed-effects regression trees + random effects included in estimation of final ensemble

R package **`pre`**: fits prediction rule ensembles
https://cran.r-project.org/package=pre

R package **`premixed`**: fits mixed-effects PREs
https://github.com/marjoleinF/premixed

Application to real data

- Global Adult Tobacco Survey: survey in low- and middle-income countries on tobacco consumption (World Health Organization, 2011).
- Data from Indonesia (N = 709 complete cases)
- Outcome: smoking status (0=non-smoker, 1=smoker)
- 5 predictors:
  - age, gender, level of education
  - belief that smoking causes illness for oneself & people around
- Level-2 unit: multistage sampling units used to produce nationally representative sample

GATS data: Results

	AUC M (SE)	# terms M	# vars M
pre	.9243 (.006)	37.7	4.7
pre_cs	.9233 (.006)	49.4	5.0
pre_m_part	.9305 (.006)	40.8	5.0
pre_m_full	.9309 (.006)	28.1	4.7
glmertree	.9089 (.016)	-	-
glmer	.9200 (.007)	-	-

Note: All values estimated based on 10-fold CV.

Simulation

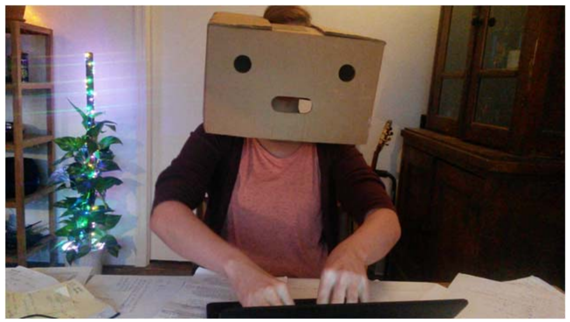
Aim: Compare accuracy, complexity and computation time

Prediction rule ensembles:

- **pre:** ignores multilevel structure
- **pre_cs**: cluster-level sampling
- **pre_m_part:** rules estimated as in pre + random effects included in estimation of final ensemble
- **pre_m_full**: estimate rules with mixed-effects regression trees + random effects included in estimation of final ensemble

Other methods:

- **lme**: linear mixed-effects model
- **lmetree:** mixed-effects regression tree
- **forest**: random forest



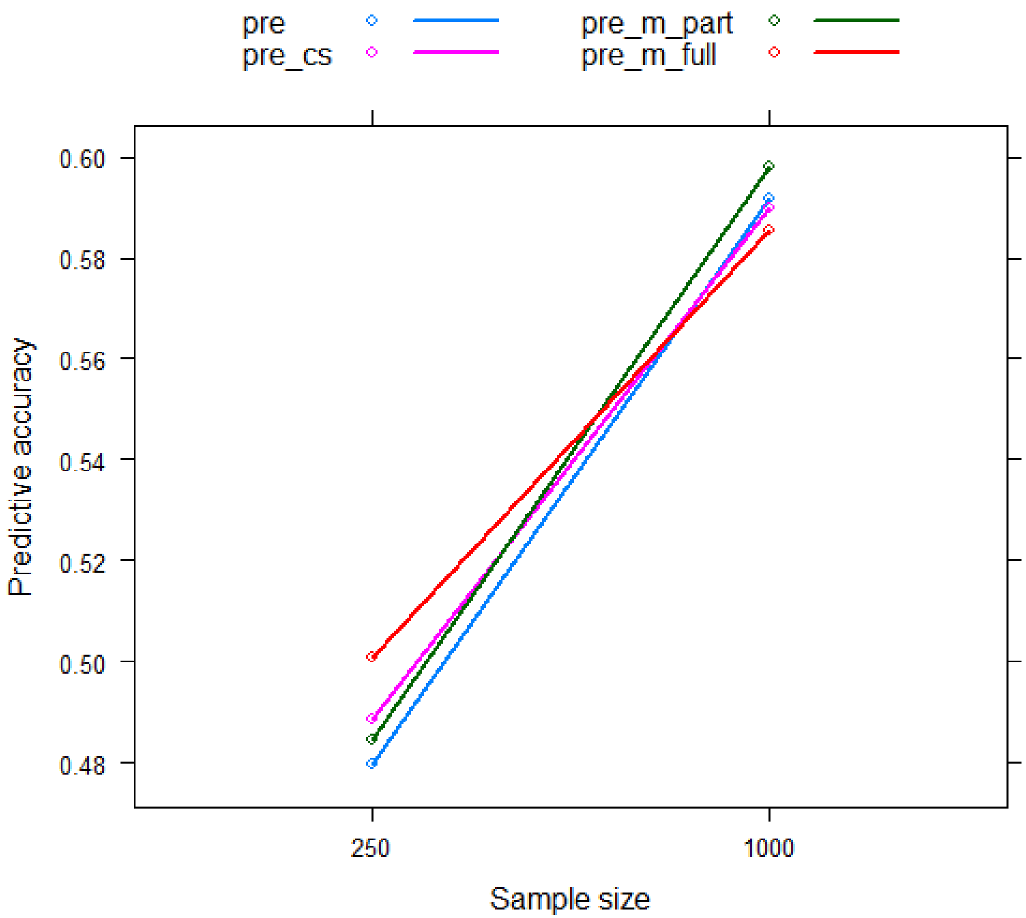
Simulation: Data-generation

True model:	Linear terms Rules Rules + linear terms Tree
Sample size	Small (250) Large (1000)
# predictor variables	Small (20) Large (50)
# of clusters	Small (10) Large (50)
variance of random effects	Small (ICC≈.1) Large (ICC ≈.2)

Preliminary simulation results

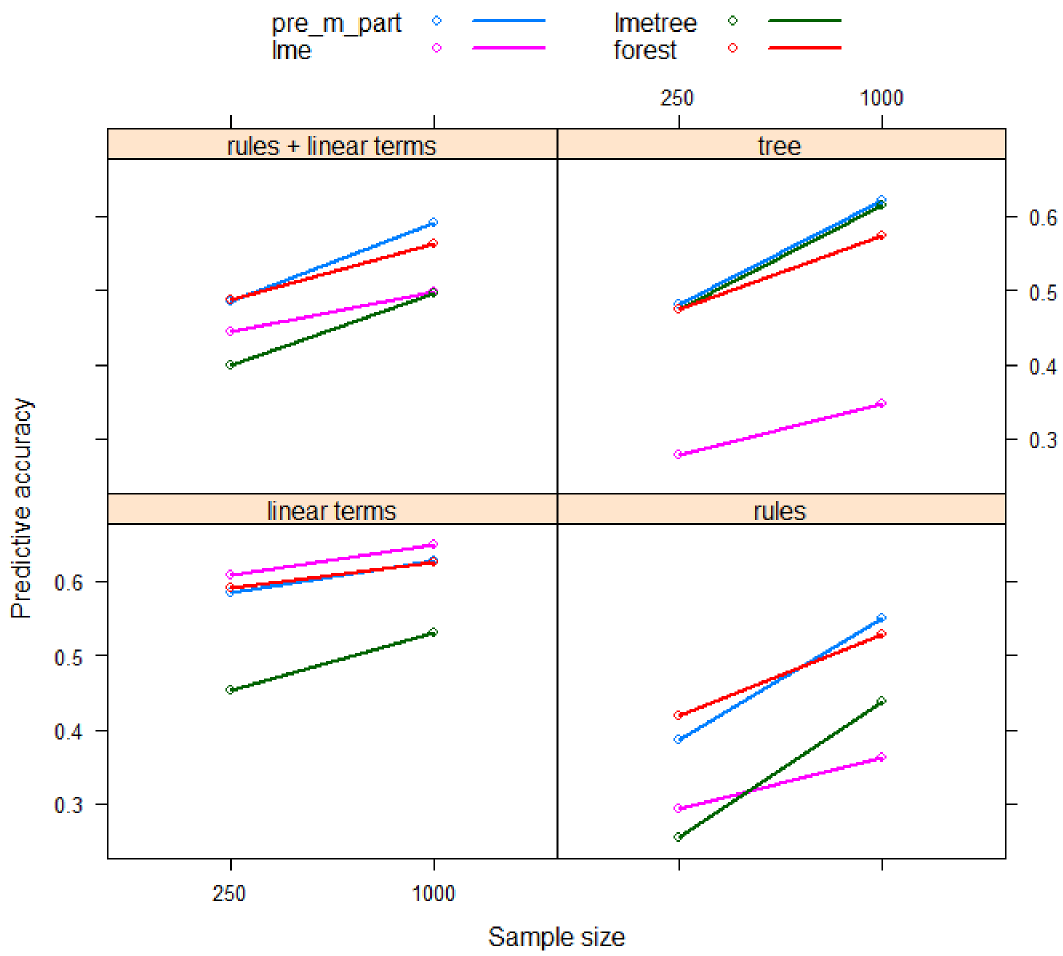
Method	Accuracy* M (SD)
pre	.536 (.090)
pre_cs	.539 (.083)
pre_m_part	.541 (.092)
pre_m_full	.543 (.081)

* Correlation between observed and predicted y in test data

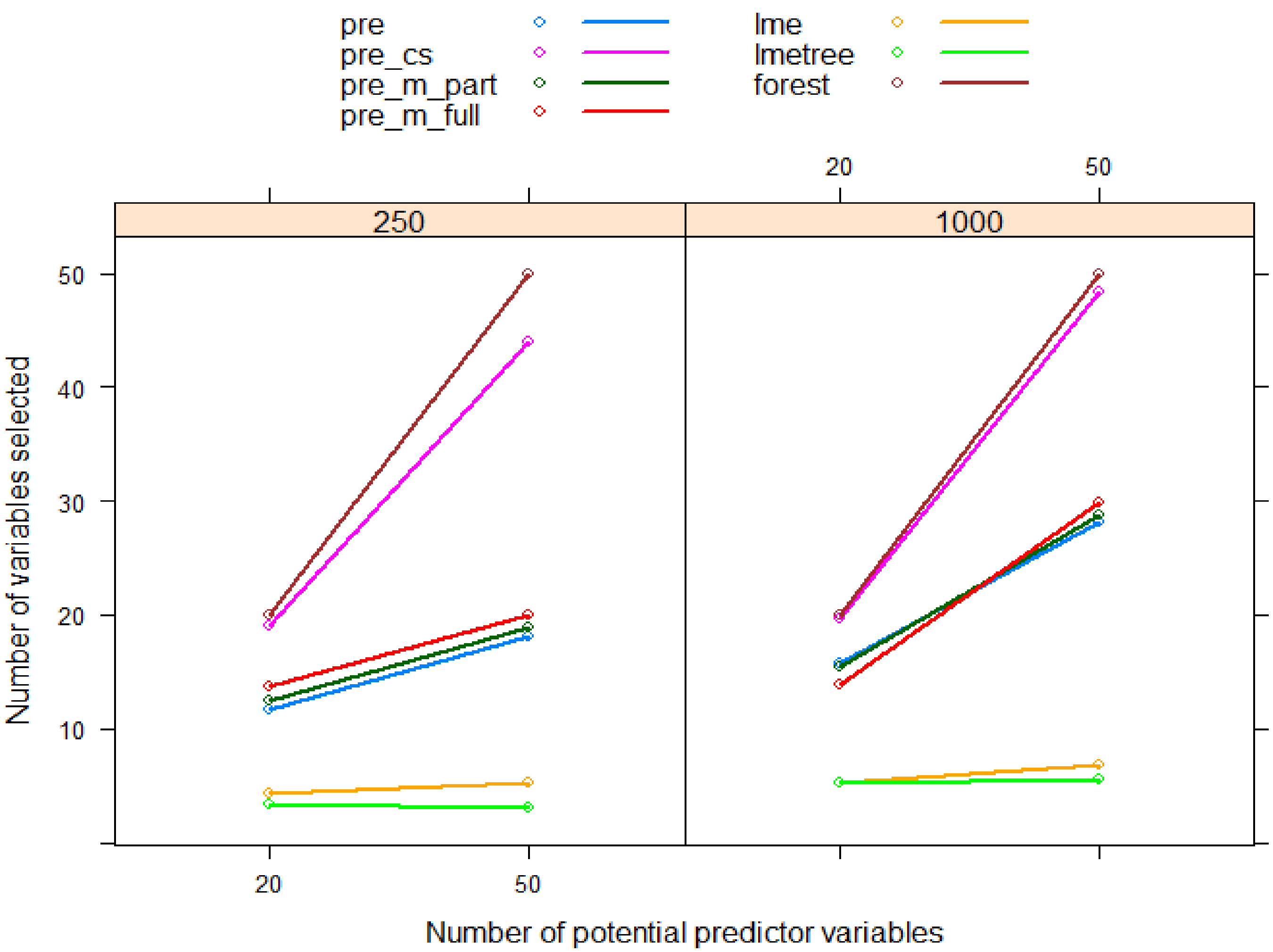


Preliminary simulation results

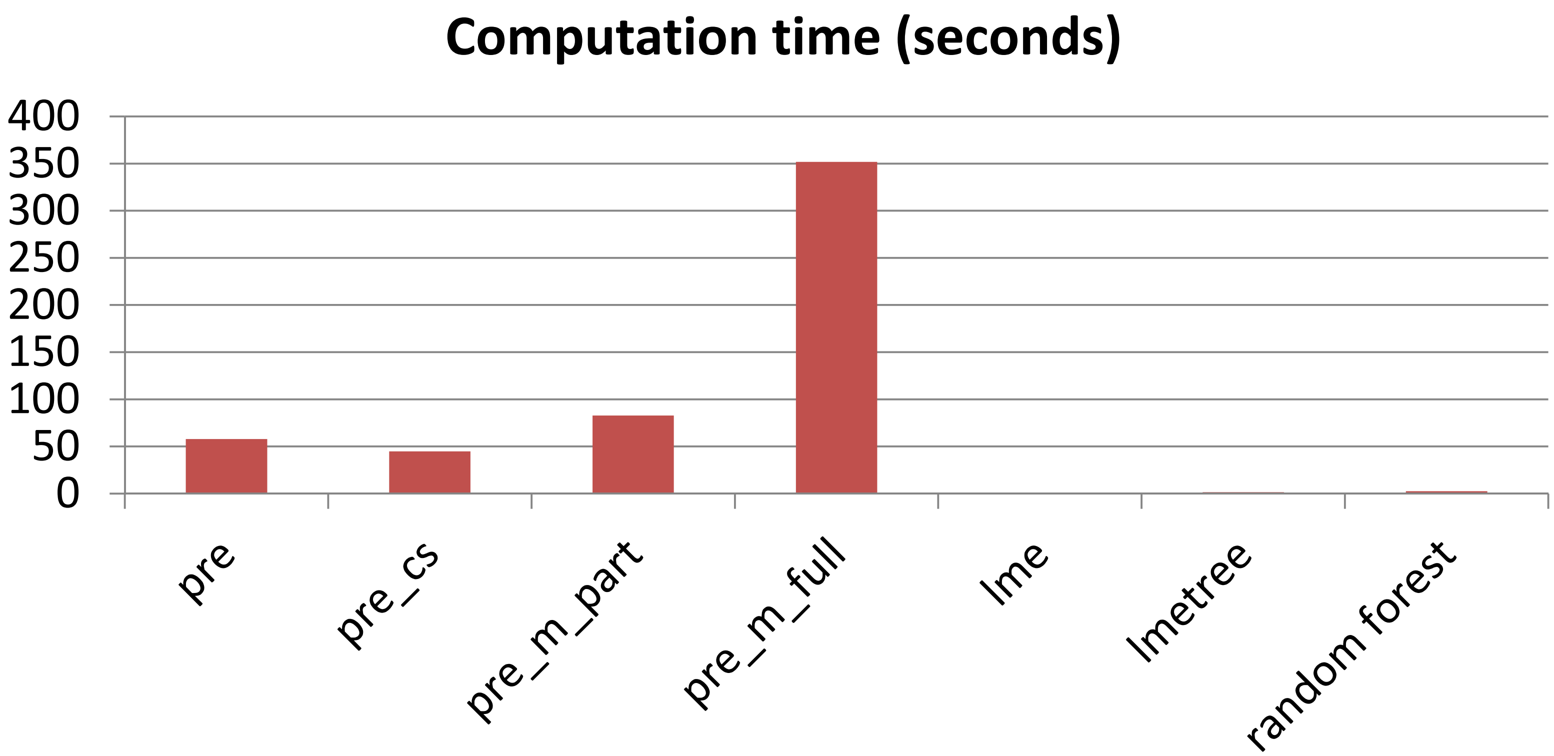
Method	Accuracy* M (SD)
pre_m_full	.543 (.081)
random forest	.533 (.080)
lmetree	.458 (.115)
lme	.435 (.140)
* Correlation between observed and predicted y in test data	



Preliminary simulation results



Preliminary simulation results



Discussion

PREs provide predictions as accurate as random forests

PREs somewhat more accurate if random effects are estimated

- Similar to findings on random forests of Hajjem et al., (2014)
- But: computationally very expensive
  - To do: Improve computation speed
- But: complexity not reduced
  - To do: Better selection algorithms than lasso (e.g., horseshoe estimator)?
- Random and fixed effects were independent in simulation, effect may be stronger if random and fixed effects are correlated
  - To do: Determine appropriate starting values if fixed and random effects correlated, and/or when ICC > .20

Cluster-level sampling may very slightly improve accuracy, but increases complexity substantially

- Similarly, Karpievitch et al. (2009) found no effect of cluster-level sampling for random forests
- To do: Why increased complexity with cluster-level sampling?

Thank you for your attention!

References

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*.

Friedman, J. H., & Popescu, B. E. (2008). Predictive learning via rule ensembles. *The Annals of Applied Statistics, 916-954.*

Hajjem, A., Bellavance, F., & Larocque, D. (2014). Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation*, *84*(6), 1313-1328.

Karpievitch, Y. V., Hill, E. G., Leclerc, A. P., Dabney, A. R., & Almeida, J. S. (2009). An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PloS one*, *4*(9), e7087.

m.fokkema@fsw.leidenuniv.nl