Point-by-point response to the reviewers' comments

Heidi Seibold

September 27, 2018

We thank the reviewers for the thorough evaluation of our manuscript and the useful and positive feedback. All comments have been addressed and the manuscript has been revised accordingly. Sections in the manuscript where text was added/changed were marked in blue (except for typo fixes).

Reviewer A

The Paper:

- Is the title of the paper descriptive and objective? : Yes

Comments (optional)::

- Does the Abstract give an indication of the software's functionality, and where it would be used?: No

Comments (optional): It should be added that the package is restricted to the comparison of two treatment conditions.

We added this information to the abstract.

- Do the keywords enable a reader to search for the software?: Yes

Comments (optional):

- Does the Introduction give enough background information to understand the context of the software's development and use?: No

Comments (optional): The introduction claims that besides data from randomised controlled trials the package is suitable for "any other two group comparison". This is misleading because it may give the impression that data sets from observational studies are also appropriate. However for the latter type of studies one may draw the wrong conclusions based on the analysis with pmtree. One might think that the treatment is only effective in one subgroup, but this is might be due to the non-random assignment of the persons to the group. The authors only make one statement about this when they explain the code with the mathematics exam example. However, the danger of misinterpreting the output when the groups are not randomly assigned should be more explicitly addressed in the paper.

We added this information to the abstract as well as the introduction. We are currently working on allowing for inverse probability weighting, yet at this point it can not be done with the package. It can so far only be done with the weights argument in the model-based trees in the partykit package (mob(), lmtree(), etc.).

- Does the Implementation and Architecture section give enough information to get an idea of how the software is designed, and any constraints that may be placed on its use?: No

Comments (optional): In this section the authors state that "Users can take a simple model they know as basis". However, the constraints of this model should be explained better. For example, the base model should always have a single dichotomous grouping variable. It is now suggested that the grouping variable may also have more categories or even may be continuous. In the help file of pmtree, the description reads as "input a parametric model and get a forest". This description is not right.

Thank you for the helpful hints. We added the information on which types of models can be used currently and that we are at the moment restricting to binary covariates. Also we updated the description. Thank you for finding the mistake.

- Does the Quality Control section adequately explain how the software results can be trusted?: Yes

Comments (optional):

- Does the Reuse section provide concrete and useful suggestions for reuse of the software, for instance: other potential applications, ways of extending or modifying the software, integration with other software?: Yes

Comments (optional)::

- Are figures and diagrams used to enhance the description? Are they clear and meaningful?: Yes

Comments (optional):: However, the model tests within each leaf (Figure 2) might be too optimistic. A note should be added that the tests of the significance of the group effect within the subgroups is a pseudo test (due to the exploratory nature of the partitioning algorithm).

You are absolutely correct. Thanks for pointing this out. We added a comment on this.

- Do you believe that another researcher could take the software and use it, or take the software and build on it?: Yes

Comments (optional):: Please add all the packages need to be installed to run the code. For the example data tje package psychopools is needed.

Thanks for pointing this out. We now mention all the packages to be installed. We also added references to the packages.

The software:

- Is the software in a suitable repository? (see http://openresearchsoftware.metajnl.com/about/editorialPolicies#custom-0 for more information) : Yes

- Does the software have a suitable open licence? (see http://openresearchsoftware.metajnl.com/faq/#q5 for more information): Yes

Comments (optional)::

- If the Archive section is filled out, is the link in the form of a persistent identifier, e.g. a DOI? Can you download the software from this link?: Yes

Comments (optional)::

- If the Code repository section is filled out, does the identifier link to the appropriate place to download the source code? Can you download the source code from this link?: Yes

- Is the software license included in the software in the repository? Is it included in the source code?: Yes

Comments (optional)::

- Is sample input and output data provided with the software?: Yes

Comments (optional)::

- Is the code adequately documented? Can a reader understand how to build/deploy/install/run the software, and identify whether the software is operating as expected?: No

Comments (optional):: I would add also an example how to use zformula.

We added an example on how to use the zformula in the documentation of pmtree(). Thank you for pointing out the need for better documentation. We are always happy to improve it to make the package easy to use.

- Does the software run on the systems specified? (if you do not have access to a system with the prerequisite requirements, let us know): Yes

Comments (optional):: I tested the software on Mac and Windows. I do not have Linux.

This is optimal as we tested on Linux (Ubuntu and Debian) and Windows.

- Is it obvious what the support mechanisms for the software are?: Yes

Comments (optional)::

Summary comments to the author(s):

- Please provide a list of your recommendations, indicating which are compulsory for acceptance : Compulsory adjustments:

1. The function pmtree() should be improved. After testing, it seems that the output of pmtree() is not right. The numbers in the leaves of Figure 2 are not right. Given the split points on the variable tests the numbers should be in node 3: n=69, in node 4, n=411, in node 6 n=86, in node 7, n= 86. Consequently, the estimates of the linear model are different. For example in node 6:

```
bmod_mathn6 <- lm(pcorrect ~ group, data = MathExam[MathExam$tests<=92.308&MathExam$tests>84.615,])
cbind(estimate = coef(bmod_mathn6), confint(bmod_mathn6))
This gives:              estimate    2.5 %    97.5 %
(Intercept) 65.19546  60.41644 69.974476
group2       -4.78822 -10.82954  1.253104
```

Thank you for finding this issue. The reason for the differences you encountered are due to the rounding of cut points in the plot output. So rather than an error in the algorithm this is a rounding issue, which we do for estetical purposes in the plot. We have currently no good solution to automatically avoid the issue. To resolve the issue the following code could be used:

```
edge <- function(obj) edge_simple(obj, digits = 4)
class(edge) <- "grapcon_generator"
tr_math <- pmtree(bmod_math, control = ctree_control(maxdepth = 2))
plot(tr_math, terminal_panel = node_pmterminal(tr_math,
  plotfun = lm_plot), edge_panel = edge)
```

For the sake of simplicity of the code shown we would like to stick to the current version, but included a note in the manuscript. Also we print the tree to the console to allow inspection of the cupoints in unrounded form.

2. The typos should be corrected: -characteristicum -wrote the exam: change into: took the exam -estimtates -theses models

Thank you for finding the typos. We fixed all of them.

- Please list any comments that are optional but would improve the quality or the reusability of the software::

Reviewer B:

The Paper:

- Is the title of the paper descriptive and objective? : Yes

Comments (optional):: The title of the paper is well-suited for the software.

- Does the Abstract give an indication of the software's functionality, and where it would be used?: Yes

Comments (optional):

- Do the keywords enable a reader to search for the software?: Yes

Comments (optional):

- Does the Introduction give enough background information to understand the context of the software's development and use?: Yes

Comments (optional):

- Does the Implementation and Architecture section give enough information to get an idea of how the software is designed, and any constraints that may be placed on its use?: Yes

Comments (optional):

- Does the Quality Control section adequately explain how the software results can be trusted?: Yes

Comments (optional):

- Does the Reuse section provide concrete and useful suggestions for reuse of the software, for instance: other potential applications, ways of extending or modifying the software, integration with other software?: Yes

Comments (optional)::

- Are figures and diagrams used to enhance the description? Are they clear and meaningful?: Yes

Comments (optional):: There are very nice and appealing illustrations that aid in understanding the context. I am wondering whether the plots are termed "bee plots" (as stated in the paper) or rather "beeswarm plots".

You are absolutely correct. We changed to the correct term.

Further, I am not sure whether I fully understand the dependence plots. A clarification would be very helpful; please see my comment below.

We added more explanations in the manuscript. See also our answer below.

- Do you believe that another researcher could take the software and use it, or take the software and build on it?: Yes

Comments (optional):: Yes, it is definitely very useful, is easily and comfortably applicable and has relevance to many disciplines.

The software:

- Is the software in a suitable repository? (see http://openresearchsoftware.metajnl.com/about/editorialPolicies#custom-0 for more information) : Yes

- Does the software have a suitable open licence? (see http://openresearchsoftware.metajnl.com/faq/#q5 for more information): Yes

Comments (optional)::

- If the Archive section is filled out, is the link in the form of a persistent identifier, e.g. a DOI? Can you download the software from this link?: Yes

Comments (optional)::

- If the Code repository section is filled out, does the identifier link to the appropriate place to download the source code? Can you download the source code from this link?: Yes

- Is the software license included in the software in the repository? Is it included in the source code?: Yes

Comments (optional)::

- Is sample input and output data provided with the software?: No

Comments (optional):: No, but example data can be easily retrieved from a different R package (psychotools). This is also well described in the article, so I do not see any need for action.

- Is the code adequately documented? Can a reader understand how to build/deploy/install/run the software, and identify whether the software is operating as expected?: Yes

Comments (optional)::

- Does the software run on the systems specified? (if you do not have access to a system with the prerequisite requirements, let us know): Yes

Comments (optional)::

- Is it obvious what the support mechanisms for the software are?: Yes

Comments (optional)::

Summary comments to the author(s):

- Please provide a list of your recommendations, indicating which are compulsory for acceptance : I really much enjoyed reading the article. The R package implements methods highly useful to a variety of different areas and I expect the paper and software to attract many readers. The description and illustrations are nice and well understandable. Overall, I do not have any major concerns, but only the two following minor concerns.

We are happy that you liked the article and the package.

(1) Although well written in general, there are several typos in the manuscript (e.g. "datam"," y-axsis", "successfull", "as all variable not given..", "asses", "anwers", "estimtates", "theses models"). The authors should carefully go through the manuscript again (including the R code –> "estmated") and make sure to correct all typos. I'd also recommend writing "do not" instead of "don't" since its more formal.

Thank you. We fixed all the typos and ran another round of spell checks.

(2) Even after having read the section in Seibold et al. (2017) I still have problems in understanding the concept of the "dependence plots". From my understanding these plots are not the same as the well-known "partial dependence plots", since the plots the authors show do not show the net effect of the covariate (ie controlled for other covariates). E.g. we might consider gender to play a major role (as suggested by the figure), although when controlling for the no. of attempts, there might be no difference between males and females anymore (obviously from the illustrations gender and no. of attempts seem to correlate). Is my understanding correct? I recommend adding an explanation regarding this issue since I think it is very important aspect. A short interpretation of (at least one of the) example dependence plots would be very helpful for clarification as well.

We added some clarification on this matter in the manuscript. The dependence plots shown are not the same as the classical partial dependence plots. Each dot in the plot relates to one patient. This way we can even show interactions of student characteristics (as in Figure 4b).

Besides the issues mentioned, I do not have any compulsory recommendations.

Please list any comments that are optional but would improve the quality or the reusability of the software:: Even though the authors correctly state at some place in the manuscript that causal interpretations cannot be done here due to the design, I'd recommend to completely avoid phrasings like "To investigate whether the exam was fair, . . . ".

We agree and changed the wording.

I can understand that the users restricted the tree depth to two layers for simplicity; I'd suggest adding an explanation on the maxdepth parameter to make clear that even more granular subgroups might be found.

Thank you for this comment. We added the information to the manuscript.

In the light of the title and the strong relevance to personalized medicine, I was somewhat astonished to not see an example from the medical field. Although the example given is well understandable, the authors may want to consider if an example from the medical field would possibly attract more readers. But the current example is also very nice.

We decided to use an example from a different field to show the relevance for a broader audience.