Authors response to the reviewer

Brandon M. Greenwell & Bradley C. Boehmke

1/19/2020

We wish to express our thanks to the anonymous reviewer who provided a thoughtful analysis of our work. We believe the changes we made, based on their suggestions, have resulted in a much better presentation of the **vip** package, and the general topic of *variable importance*. Below we outline the changes we've made in regards to the reviewer's comments.

Summary of overall changes

- Based on the reviewer's thoughtful suggestions, we've expanded the discussion in various sections to include drawbacks, additional R packages, etc.
- The **vip** package was updated on CRAN, and some changes were made to the code examples to reflect changes in argument names, etc. None of the output has changed.
- We've slightly restructured the terminology of the section on PDP- and ICE-based VI scores. We've recently found that this method falls as a special case of the *feature importance ranking measure* discussed in Alexander Zien, et al. (2009). The only difference in this section is the terminology, additional references, and changes in the code to reflect changes in the package (which is now on CRAN).
- Since our initial submission, we've expanded the functionality of **vip** to include Shapley-based importance scores (these are important enough that the methodology should've been discussed more in depth anyway). We figured this was important enough to be added to the paper. This is a minor section (with relevant references) with one example. If the reviewer, or editors, find this to be unnecessary, we're happy to remove it (but we strongly think it's a useful contribution).

Response to reviewer

Reviewer: The authors miss some important literature I think. most especially the functional ANOVA as developed by Giles Hooker. this is important work (though he doesn't have a working software implementation last I checked) which discusses what I think are some of the shortcomings of this work.

Authors: Thank you for pointing us to Hooker's work. We actually found his more recent work, "Please Stop Permuting Features: An Explanation and Alternatives", to be more relevant. Nonetheless, we still found Hooker's general functional ANOVA work relatable to the FIRM-based approach described in our paper and have included a brief discussion and references to both in our paper.

Reviewer: These methods don't generally do well when the function that maps the features to the outcome, as learned by whatever ML model used, is actually high dimensional, non-additive, etc. this I think warrants a bunch of big caveats as this is the situation where ml really does well relative to other methods. I'd love to see a section where you show how to break all these methods.

Authors: Agreed, and briefly discussed in our new section on "Drawbacks of existing methods" (these issues are also pointed to in Hooker's general functional ANOVA work).

Reviewer: Another thing I think that is really missing is a discussion of the concept of faithfulness (to the data generating process? the learning process? the fitted model?) which speaks to what the purpose of doing this sort of analysis is.

Authors: We think of "faithfulness" as it pertains to prediction explanations and to what degree they accurately represent the underlying "truth" of the data generating process, or "true" underlying mechanisms driving the outcomes. To us, this seems to be a two step process:

(1) Understand what features are driving our model's predictions and how it is doing so across the input space.
(2) Determining how well this relates to the "true" underlying mechanisms.

Our paper addresses (1). We even state "While many of the procedures discussed in this paper apply to any model that makes predictions, it should be noted these methods heavily depend on the accuracy and usefulness of the fitted model; hence, unimportant features may appear relatively important (albeit not predictive) in comparison to the other included features."

We then go on to state: "For this reason, we stress the usefulness of understanding the scale on which VI scores are calculated and take that into account when assessing the importance of each feature and communicating the results to others."

Although we believe (2) distracts too much from the main points of this paper, we have briefly expanded this part of the discussion (e.g., in the discussion on drawbacks of existing methods) and added references to more in-depth treatments.

Reviewer: Related, the author(s) don't really ever go into detail about what 'interpretable' or 'important' mean: which is related to the question of faithfulness. These are important definitions that I don't think can be safely ignored, especially without much theoretical literature to reference.

Authors: Agreed, there is not much theoretical literature to reference on these terms, which are rather difficult to assign formal definitions to. Nonetheless, we've added some discussion (and footnotes) providing our own working definition of both, as well as some references to more formal (and recent) treatments. We especially appreciate the recent work by Patrick Hall and Christoph Molnar.

Reviewer: I hasten to add that the authors missed some of the theoretical literature on the performance of some of these methods. I don't recall of the top of my head but I know there has been work done on the RF OOB permutation methods by the authors of the party package, as well as one of the authors of the PDP module of scikit-learn. It is not important but they also missed my work published here (mmpf).

Authors: This is a good point. We are very aware of the work of Carolin Strobl (w/ others) on conditional permutation-based variable importance in RFs (which was always built into **vip**). We mistakenly omitted such work in our initial draft but have added it in the relevant sections. We'd also like to thank the reviewer for pointing us to their own work on **mmpf**, which we were previously unaware of. We included it in our discussion and also found it useful to add to our benchmark of permutation-based methods.

Thank you again for your time and response.

Best,

Brandon G. & Bradley B.