

Response to review

Cheem Exploring Local Explanations of Nonlinear Models Using Animated Linear Projections
November 10, 2023

Thank you to the editor and reviewers for your helpful feedback on our paper. We have revisited the manuscript and to address the reviewer's comments:

We respond the reviewer comments inline in blue, and large changes in the paper are also marked in blue.

Changes made also reflect a change in the R package, with the cheem v0.4.0 is now hosted on CRAN as of 11/08/2023.

*Thank you,
Nicholas Spyrison,
Corresponding author*

COMMENTS FOR THE AUTHOR:

The authors should update their references using the Computational Statistics citation style.

We now follow the guidelines provided at the journal web site and use the Springer Nature template as suggested.

Reviewer #1: This paper proposes an interactive visualization tool and its implementation as an R package to intuitively understand LVA in non-linear models.

General Comments

You have not clearly stated the usefulness of the radial tour; please explain the advantage of using the SHAP values of PI and CI as the contribution of 1D projection compared to drawing a static density plot or scatterplot. Is it possible to correctly evaluate the sensitivity of variables by 1d projection for nonlinear models? Also, in the Case Studies in Section 5, please provide an explanation in relation to the Radial tour operation.

Section 3 introduced the radial tour and provided a simple explanation of how it is generally used for understanding variable importance. Section 4.2 connects this to the use of SHAP

values for explaining the prediction for an individual observation. We have added a paragraph with the general thinking.

SHAP values are effectively locally linear estimates of variable importance for a prediction. Thus, yes, in the local neighbourhood of an observation, testing the sensitivity of variables for the prediction can be done using 1D projections. We have added a sentence to this effect.

We have also added some additional wording in Section 5 case studies.

The results are very different between the examples provided in the shiny application in the R cheem package and the examples presented in the paper. Please make sure to use random number seeds to achieve similar results. In particular, in the Case Studies of classification, there is no misclassification in the examples in the cheem package, and the misclassification is not indicated by the red circle as shown in the paper.

We have revisited the attached models and objects and verified that classification models misclassify some observations for the purposes of discussion.

The panel that is treeshap in the global view in the paper is attribution in the cheem package. Please unify either one.

We have clarified what the panels depict in the application and figures 3 and 4. This will meaningfully vary from other figures in this paper. As we use specific models and attributions to illustrate, while the package allows for the general case.

In the global view in the paper, the model panel has jitter processing, but the cheem package does not.

Classification models do use jitter to avoid occlusion of the observations. While regression models have more continuous distributions. This has been clarified in the panel titles.

Discussion is just a Conclusion. For example, please supplement any discussion on the following.

- Abstract and Introduction are described in the context of XAI, but the tools proposed in this paper are not likely to be able to handle large data sets.

Done, changes are highlighted in blue.

Concerning the convenience of the package: the application by the cheem package outputs little information (e.g., axis information), which makes it difficult to use intuitively.

We have revisited the panel strips to better describe what is being viewed.

Individual Comments

(p.7 l.18) Figures 3b, 4a -> 3c, 4b?

You are correct, this has been fixed.

(p.8 l.7) Is the SHAP value of the CI taken into account in the radial tour?

In the explanations of the radial tour we have more clearly stated that it starts from the PI SHAP values. When changing the projection values one could make it closer to those of the CI, which is done in some examples.

(p.8 l.8) "The PI is ... global view" : The shiny app in the cheem package does not support interactive selection by clicking in the global view. (cheem version 0.3.0 on Windows 11 / Google Chrome)

We have clarified the text from selection of the PI to click and drag highlighting discussed above and below this point. We have confirmed that plotly interaction is listed as an import. And validated the interaction (cheem version 0.3.0 on Windows 10 / Google Chrome).

(p.10 Fig4) Residual plot is not displayed in cheem version 0.3.0 on Windows 11 / Google Chrome.

We have confirmed that residual plots do display for regression models, while classification models give more of a visual confusion matrix display. (cheem version 0.3.0 on Windows 10 / Google Chrome).

(p.11 l.5) Figure 3a & b -> 4a & b?

We have corrected the manuscript, thank you.

(p.11 l.7) Figure 3d -> 4d?

We have corrected the manuscript, thank you.

(p.21 l.10) <https://nspyrison.github.io/cheem/> is not available.

The packagedown site has been recovered, updated version has been submitted to CRAN.

(references) There are some characters like "???".

Fixed

Reviewer #2: This is a very interesting paper.

I was able to run your software using publicly available packages.

Unfortunately, I could not play videos shown at the following URLs.

<https://vimeo.com/666431172>

<https://vimeo.com/666431143>

<https://vimeo.com/666431148>

<https://vimeo.com/666431163>

<https://vimeo.com/666431134>

It would be even easier to understand if videos could be played back, so I hope this will be possible.

We cannot address this unless the error messages are provided. The videos at each of the addresses are playable by us and several colleagues who also checked. We have validated that these url's are pointing to the correct content, videos are set to public and content is marked for all audiences.

I think the visualization you propose is very clear.

However, I have the impression that it is difficult to understand, so I expect detailed explanations.

For example, detailed explanations of Figure 3 e and Figure 4 e would be helpful.

I also think that explanations of the data in Figure 3 and Figure 4 would make it even easier to understand.

See the response to reviewer 1. The new text gives more details that should make it easier to understand.

For small details, I think it is necessary to check page 5, line 38 and page 11, line 29.

Page 5, line 38, discusses the R packages related to LVA. These packages have been validated to be available on CRAN. treeshap has been updated to reflect it's availability on CRAN. Text and references have also been updated for accuracy version number at time of clarification 10/24/2023.

Page 11, line 29 discusses the hyperparameter values used in the random forest models. This sentence has been parsed into two for readability, regression and classification respectively. The values have been validated to still be accurate to the examples used.