Author's response to reviews

Title: EFS: An Ensemble Feature Selection Tool implemented as R-package and Web-Application

Authors:

Ursula Neumann (u.neumann@wz-straubing.de)

Nikita Genze (nikitagenze@yahoo.de)

Dominik Heider (d.heider@wz-straubing.de;dominik.heider@uni-due.de;dominik.heider@uni-marburg.de)

Version: 1 Date: 04 May 2017

Author’s response to reviews:

Reviewer 1:

1. First, while they walk through an example dataset, it is unclear what range of data types might be appropriate for their method. Adding a bit of discussion of what types of experimental designs and data types are supported could increase the impact of the software.

> We appreciate the reviewer’s suggestion and acknowledge that this would be helpful for the reader. We added a corresponding paragraph in the methodology section. The type of the experimental design is a binary classification, e.g., random clinical studies. The two classes have to be stored in one dataset with a class label as an additional column.

2. Finally, while the main goal of this manuscript is to describe the implementation of their approach, a bit more description of the method itself would make the manuscript be more readable as a stand along manuscript.

> We followed the suggestion of the reviewer and added a few more explanatory sentences to some of the more complex methods.

Reviewer 2:

1. Page 1, Abstract, Conclusion: The authors may consider revising the first sentence to "The EFS identifies relevant features while...".

> We followed the suggestion of the reviewer and rephrased the corresponding paragraph.
2. In the Introduction, the authors may consider adding a very brief (1-2 sentence long) general description of why feature selection is important and how it is typically used (e.g., a sensitivity analysis to reduce data dimensionality by ranking a list of features by importance and then eliminating those features that are less relevant to the data mining/machine learning question. This improves the overall performance of the model because it addresses the problem of overfitting.). This will demonstrate the outright importance of using the described methods.

> We appreciate the reviewer's suggestion and added a corresponding paragraph.

3. Page 2, under "Median": "As importance measure the p-values were used." is an awkward sentence. Do the authors mean to state "The resulting p-values are used as a measure of feature importance."?

> We followed the suggestion of the reviewer.

4. Page 2, under "Random Forest": the authors should spell out the abbreviation of area under the ROC curve (AUC) when it is first used. Also, ROC should be defined here as well.

> We followed the suggestion of the reviewer.

5. Page 2, "Ensemble learning". Words in all of the other subheadings start with capitalized letters (should this be "Ensemble Learning"?)

> We followed the suggestion of the reviewer.

6. Page 3, last paragraph under "R-Package": A period (.) is used where a comma (,) should be: Consider revising to "Due to the high computational costs of the RFs, the default selection is set to"

> We followed the suggestion of the reviewer.

7. Page 4 under "Performance evaluation by logistic regression": Receiver operating characteristic (ROC) curve should be spelled out on page 2 when the abbreviation is first used (See comment above).

> We followed the suggestion of the reviewer.

8. Page 4, under "Permutation of class variable": The authors should consider revising the last sentence to "...the AUC from the original LR model using a Student's t-Test." or some similar wording.

> We followed the suggestion of the reviewer.

9. Page 5, last paragraph under "Example": "Figure 2" is stated with a capital "F" however "figure 3" is not. Should this be "Figure 3"?
We followed the suggestion of the reviewer.

10. Page 5, last paragraph under "Example": The sentence "Low values of variances imply that the importances are stable and reliable" is awkward. The authors may consider revising this sentence to "A low value of variance implies that the importance of a feature is stable and reliable" or some similar wording.

We followed the suggestion of the reviewer.