Author's response to reviews

Title: Regression Tree construction by bootstrap: Model search for DRG-systems applied to Austrian health-data

Authors:

Thomas Grubinger (thomas.grubinger@i-med.ac.at)
Conrad Kobel (conrad.kobel@i-med.ac.at)
Karl-Peter Pfeiffer (karl-peter.pfeiffer@i-med.ac.at)

Version: 4 Date: 9 January 2010

Author's response to reviews: see over
Author’s response to reviews

Title: Regression Tree construction by bootstrap: Model search for DRG-systems applied to Austrian health-data

Authors:

Thomas Grubinger (thomas.grubinger@i-med.ac.at)
Conrad Kobel (conrad.kobel@i-med.ac.at)
Karl-Peter Pfeiffer (karl-peter.pfeiffer@i-med.ac.at)

Version: 3 Date: 9 January 2010

Author’s response to reviews: see over
Dear Ms Clark,

we thank you and the reviewers for the careful review and the constructive comments on our paper. We have revised our manuscript based on each of the reviewers’ comments as well as the editorial points and respond on them below.

Response to reviewers’ comments: Jay M. Rosenberger

1. Minor Essential Revisions

   We agree with all the reviewer’s comments in this section and have corrected them as suggested.

2. The definition of the “size” of a tree is usually, the total number of nodes (i.e. size = internal nodes + terminal nodes). While I believe that the authors intended this to be the definition in this paper as well, they use terminal nodes in the equation on page 6, and they use internal nodes in the equation on page 9. Consequently, this review was wondering whether the authors were using any alternative definition of size. In any case, I recommend that the authors define “size” clearly.

   Despite the equation on page 6 we always used the number of internal nodes as a measure of tree size or complexity. This is because they can simply be interpreted as the number of rules required to classify patients. We have now defined and motivated our measure of tree size and describe it’s relation to the number of terminal nodes and the total number of nodes. (2nd paragraph in subsection: Regression Trees)

   We also explain the formula on page 6 in terms of internal nodes.

3. The authors also use the term “complexity” and “topology” of a tree. These seem to mean the same thing, so I recommend using only one of them or acknowledge that they are the same.
A binary tree with a given number of nodes can have different forms, i.e. the internal nodes can be arranged differently. This is the definition of topology that we use in our paper.

We have now clearly defined the term “topology” in the manuscript. (Subsection: Model Search by Bootstrap)

4. Similarly, the authors also use the term “split,” which they define clearly, and “rule,” which they leave undefined, to mean the same thing. I recommend using only one of them or acknowledge that they are the same.

Depending on the context we use “split” and “rule” exchangeable. We acknowledge that they both refer to the same thing. (2nd paragraph in subsection Regression Trees)

Response to reviewers’ comments: Andrew A. Neath

Results are presented showing that the bootstrap bumping procedure achieves a solution with improved accuracy over a traditional CART algorithm. I wonder if the authors can provide some feedback on how well the bootstrap bumping procedure achieves the second goal of a medically interpretable model. Have the authors put their idea into practice with medical experts? Have medical experts been receptive to the idea? Are there difficulties in selecting a model based on medical considerations in practice? I suggest the authors add a section to the paper providing details on this aspect of their proposed solution.

During the whole development of the Austrian DRG-system medical experts have been involved in the evaluation of the resulting regression trees and many times not the statistical optimal tree has been selected due to medical expert opinion. The advantage of this algorithm is that more than one solution is presented for the discussion with medical experts. This presentation shows the possibilities which will be used in the next years for the maintenance and further development of the Austrian DRG-model. From the discussion with medical experts we know that a single “data driven” model is not always the medical correct one and different options have to be presented for medical evaluation.

We have included the information above into the ‘Conclusion’ of the paper.

Response to the editorial points

1. We recommend that you copyedit the paper to improve the style of written English.

We have copyedited the paper to improve the style of written English.

2. Please include a ‘Competing interests’ section between the Conclusions and Authors’ contributions.

There are no competing interests. We have included the ‘Competing interests’ section between the Conclusions and Authors’ contributions.
3. Please acknowledge anyone who contributed towards the study by making substantial contributions to conception, design, acquisition of data, or analysis and interpretation of data, or who was involved in drafting the manuscript or revising it critically for important intellectual content, but who does not meet the criteria for authorship. Please also include their source(s) of funding. Please also acknowledge anyone who contributed materials essential for the study. Authors should obtain permission to acknowledge from all those mentioned in the Acknowledgements. Please list the source(s) of funding for the study, for each author, and for the manuscript preparation in the acknowledgements section. Authors must describe the role of the funding body, if any, in study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; and in the decision to submit the manuscript for publication.

We have included an “acknowledgements” section.

4. Please could you clarify whether the datasets you used were publically available, or whether you obtained permission to use the data.

The data is not publically available. Permission to use the data was granted from the Bundesministerium für Gesundheit, Familie und Jugend. (Added to subsection: The DRG-Data)

Sincerely,

Thomas Grubinger, Conrad Kobel and Karl-Peter Pfeiffer