Author's response to reviews

Title: A Comparative Analysis of Multi-Level Computer-Assisted Decision Making Systems for Traumatic Brain Injuries

Authors:

Soo-Yeon Ji (jisy@vcu.edu)
Rebecca Smith (smithr@vcu.edu)
Toan Huynh (toan.huyhn@carolinashealthcare.org)
Kayvan Najarian (knajarian@vcu.edu)

Version: 3 Date: 2 June 2008

Author's response to reviews: see over
Over the last 20 years, the development of computer decision support systems has been a growing field in the research medical domain. Trauma experts must make several difficult decisions considering a large number of patient attributes, often in a short period of time. These decisions are made at different stages of care giving, based on the available patient data at the time of each decision. Computer-aided systems can provide vital assistance to clinicians in deciding both diagnosis and treatment, and so provide a higher standard of patient care. This paper focuses on current medical issues, and proposes a predictive computer-assisted decision making system for traumatic head injury (TBI) using machine learning algorithms. Machine learning provides a good solution for dealing with missing values, a critical issue when working with medical data.

Thus, the main goal of this study is to develop a reliable rule-based decision making system that provides both recommendations and outcome predictions at critical stages of care giving and compare the existing machine learning methods available for medical informatics. We also suggest that using all available variables in rule generation may not be optimal for medical applications. The inclusion of less relevant and/or reliable attributes, with regards to some predefined means of measurement, can create random correlation in some rules that are ultimately clinically meaningless. Furthermore, retaining less informative and/or highly correlated attributes increases the length and complexity of the resulting rules, and so makes them less desirable for clinical application. To support this assertion, we compare five machine learning algorithms using both the all-available-variables and significant-variables-only methods, and then generate the most reliable rules using CART and C4.5, which are chosen due to their transparency. Sensitivity and specificity are also calculated to evaluate outcome identification.
Response to Reviewer’s Report  
(A Comparative Analysis of Multi-Level Computer-Assisted Decision Making Systems for Traumatic Brain Injury)

Regarding the issues with the order of sections, the Methods and Conclusions sections have been swapped. The LaTeX template downloaded from the site had these in the incorrect order, hence the change in the second version of the manuscript.

To address the reviewer’s specific concerns:

Major concerns:

1. **The new description of “Direct MLE” is inadequate. Do the authors mean that all possible variables were entered into the model together and a single likelihood-ratio test was performed? It simply isn’t clear.**

As suggested by the reviewer, a new description of Directed MLE is provided:

   To test the significance of the individual variables, we compare a reduced model that drops one of the independent variables with a full model using log-likelihood test. The test for a significant difference between the full model and the reduced model simply uses the ratio of the maximized value of the likelihood function for the full model ($L_1$) over the maximized value of the likelihood function for the reduced model ($L_0$):

   \[ -2 \log \left( \frac{L_0}{L_1} \right) = -2 \left[ \log(L_0) - \log(L_1) \right] = -2(L_0 - L_1) \]

   If the chi-square value for this test is significant, the variable is considered to be a significant predictor. Following these tests, only the significant variables ($p$ value $\leq 0.05$) are selected. For example, the likelihood ratio tests of individual parameters show that the model without gender is not significantly different from the full model and therefore gender should be dropped. Our study refers to this method as directed maximum likelihood estimation (direct MLE). This test is performed using the SAS.

   This explanation has been integrated into the paper to better describe the direct MLE method.

2. **The authors repeatedly state that logistic regression "does not assume a linear relationship". This is incorrect. This model assumes a linear relationship between the log-odds of outcome and any nominal variable included as a covariate. For this reason it is essential to explore whether**
this assumption is reasonable before simply including nominal variables in logistic regression models.

As the reviewer suggested, the assumption (linearity) of the logistic regression is tested. Before explaining the results, it is necessarily to explain how nominal variables are formatted in our dataset. As we mentioned in our first-revision response, all nominal variables are used as individual independent variables. For instance, we have seven complications (Acute Respiratory Distress Syndrome (ARDS), Aspiration Pneumonia, Bacteremia, Coagulopathy, Intra-Abdominal Abscess, Pneumonia, and Pulmonary Embolus). Each complication is treated as a dummy variable – i.e. a categorical independent variable with two levels (Yes and No). Since we use a variety of independent variables, we present selected results for following tests. (Note that in the interests of brevity, these results have not been included in the paper; however, we wish to fully address the reviewer’s concerns on this topic)

Two methods are commonly employed to test linearity between the log-odds of outcome (logit) and predictors:

(1) Logit step test
Sub-divide the predictors into a number of categories with equal intervals, then run logistic regression again with the same dependent variable using the newly created categories. If there is linearity with the logit, the variable’s coefficient should increase (or decrease) in roughly linear steps. This result means that each sub-group and the corresponding average of logit outcome have a linear relationship.

(2) Smoothing approach
Fit a non-parametric smooth function to the empirical log-odds.

In our study, we use the logit step-test method to test linearity.

Figure 1 presents the linear relationship between log-odds of the outcome and nominal variables Myocardial Infarction and Acute Respiratory Distress Syndrome (ARDS).
Fig. 1 Linear relationship between logit and nominal variables. The x-axis represents two levels (1:Yes/ 0:No) of each dummy variable, and the y-axis represents the average of the log-odds outcome of each level.

Also, the Hosmer-Lemeshow Goodness of Fit Test was assessed using Statistical Analysis Software (SAS), with the hypothesis that our model is well fitted. According to the resulting p value (p=0.6443), we cannot reject the hypothesis; in other words, our model is well fitted.

Minor errors:

As requested, several typos have been corrected and certain sentences have been reworded. We appreciate the reviewer’s diligence in locating these errors. References have also been fixed.

To address the reviewer’s other comments:

- The reference to Fu has been removed, on the grounds that it is not considered relevant to this paper.

- Regarding the Hasford paper: though we respect the reviewer’s comments, we assert that the choice was not selective. CART is in widespread use in the medical field - a brief literature review found over 1000 papers using them - because of the reasons suggested in our paper. The raw accuracy of LR and CART is comparable; however, LR lacks the other advantages of CART, in particular the ease of understanding for medical professionals. Doctors are more willing to trust in rules where they can understand the reasoning. See also the following papers:
  - Tsien, Fraser, Long, Kennedy: Using Classification Tree and Logistic Regression Methods to Diagnose Myocardial Infarction, pp. 493-497, MEDINFO98.

• Regarding the conclusions paragraph: the novelty of this paper is in applying logistic regression for significant variable extraction prior to actual rule extraction. Though we do use a single dataset, it is sourced from the second largest healthcare system in the US, and one of the largest trauma centers. Accordingly, since this dataset was not biased towards any patient group or outcome, we believe these results are sufficiently generalizable to other trauma datasets. However, we have added a sentence to the conclusions paragraph stating that our method is specifically effective in predicting patient survival and rehab/home outcome.

• We accept the reviewer’s comment regarding the number of cases; originally we intended this to refer to an approximate number. We have changed this to an exact number in the paper.

• A reference has been added to provide further information on the ID3 algorithm (a section by Ross Quinlan in the book Machine Learning, An Artificial Intelligence Approach; we found this was easier to access than his 1979 work). We believe an explanation of the algorithm within this specific manuscript is unnecessary, as the algorithm is widely known within the machine learning field.

• As requested, the definition of cross-validation has been edited for length.

• The term Po(R) has been removed, as it is not used in any equation.

• As requested, the meaning of “sgn” has been clarified (it refers to the signum function: -1 if the argument is less than 0, +1 if it is greater than 0, and 0 if it is equal to 0).

• The definition of estimate of accuracy has been moved to the preceding paragraph, as suggested by the reviewer.

• As suggested, the explanation of logistic regression has been replaced with a brief description, supplemented with a textbook reference (specifically, Chapter 1 of Hosmer and Lemeshow’s Applied Logistic Regression, which provides a suitable overview of the approach).
Regarding the performance of logistic regression vs. other methods: the reviewer's comments are justified when considering raw accuracy. However, as mentioned previously, there are numerous other benefits of using CART for medical applications. Specifically, rules offer transparent reasoning that both appeals to physicians and can be applied to future cases. Thus, we use decision tree algorithms to generate reliable rules in our study.

To address the reviewer's comments regarding the tables:
- The number of decimal places has been reduced to 1, and is consistent across all tables, as suggested.
- The example BMC LaTeX template does suggest using a table description above the example table. However, we have tried to edit down the text we used, and where possible incorporate text into each table (without overrunning the allotted space and negatively affecting the formatting).
- Previously, the body of the paper contained definitions of terms used in the tables. As suggested by the reviewer, these have been moved to the table description sections.
- The range of GCS values has been corrected.
- The reviewer is correct in the comments regarding the total of rehab and home cases. This is not the same as the number of surviving cases, because patients who were transported to other hospitals were removed from the set (as we do not know the final outcome in these cases). This has been clarified in the description for Table 2.
- As requested, “ICU stay > 2 days” is now used.
- Cells no longer contain question marks.
- Other typos have been fixed.
- As suggested, we have removed sentences starting “This table explains”, with the exception of the description for Table 6, as we believe it is important to state which methods are being compared, and which dataset each prediction concerns.
- Accuracy is still abbreviated in the column header. Due to the length of the generated rules, this abbreviation was necessary in order to fit the table within the allotted space.