ORIGINAL PAPER - SUPPLEMENTARY MATERIAL

The use and applicability of machine learning algorithms in predicting the surgical outcome for patients with benign prostatic enlargement. Which model to use?

Panagiotis Mourmouris [1], Lazaros Tzelves , Georgios Feretzakis [2, 3], Dimitris Kalles [2], Ioannis Manolitsis [1], Marinos Berdempes [1], Ioannis Varkarakis [1], Andreas Skolarikos [1]

[1] 2nd Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens, Greece;
[2] School of Science and Technology, Hellenic Open University, 26335 Patras, Greece;
[3] Department of Quality Control, Research and Continuing Education, Sismanogleio General Hospital, 15126 Marousi, Greece.

APPENDIX A

In what follows, assume$y1$, $y2$, …$yn$to be the predicted values on test instances and $y1$, $y2$,..,$yn$to be the actual values.

a) The correlation coefficient measures the correlation and dependence between observed values. The values range between -1 and 1, where 0 indicates no correlation, 1 a very strong positive correlation, and -1 a very strong negative correlation. The correlation coefficient measures the statistical correlation between y's and y's.

$ryy=i=1n(yi-y)(yi-y)i=1n(yi-y)2i=1n(yi-y)2$
where$y=1ni=1nyi$ and $y=i=1nyi$

b) *Mean Absolute Error* (MAE) – this is an alternative calculated by the average of the individual errors without taking account of their sign. The mean-squared error tends to exaggerate the effect of outliers – instances whose prediction error is larger than the others – but the absolute error does not have this effect: all sizes of error are treated evenly according to their magnitude.
$MAE=i=1nyi-yin$

c) Root Mean-Squared Error (RMSE) is the standard deviation of the prediction errors. The sensitivity of the RMSE to outliers is the most common concern with the use of this metric since the RMSE penalizes large errors while the MAE gives the same weight to all errors (1).

$RMSE=i=1nyi-yi2n$

Concerning the values of the MAE and RMSE, the smaller their values, the better the model's performance, since both metrics are proportional to the difference between the actual and predicted values.

APPENDIX B

1) Linear Regression is a common statistical approach for constructing a linear model (function) predicting a value of the variable while knowing the values of the other variables. It uses the least mean square method to adjust the parameters of the linear model.
2) Multilayer Perceptron is performed on multi-layered neural networks, usually interconnected in a feed-forward manner, where each neuron on all layers, except the output one, feeds into the subsequent layer's neurons. The network parameters can also be monitored and modified during training time. Such networks are usually trained using an error back-propagation mechanism where the observed differences between actual and expected outputs help drive the adjustment of the weights (the connections between neurons).
3) SMOreg implements the support vector machine for regression (2). The parameters can be learned using various algorithms. The default algorithm (RegOptimizer) is the RegSMOImproved (3).
4) Instance-Based Learning (k-Nearest Neighbors) (lazy.IBk) Instance-based learning approaches (4), such as the *k-nearest neighbors* (kNN) algorithm, adopt a straightforward approach to estimate real or discrete-valued target functions (5). Predicting the output of a new input vector involves collecting and aggregating outputs from similar instances

No conflict of interest declared.

(called "*neighbors*") from the saved training data. Unlike many other techniques that create only one local approximation to the target function, an important advantage of instance-based algorithms is that the model can build a new approximation to the target function for each new query instance. This gives instance-based algorithms the ability to capture very complicated relationships between attributes and outputs. If the target variable depends only on a few of the attributes, this can cause very similar instances to be predicted at a large distance (6). In our experiments, the value of k was set to 4.

5) Meta-learning method "*Bagging*" consists of aggregating results of n models generated on the basis of *n* bootstrap sets. The bootstrap sets are generated using feature selection or random drawing with substitution from the original dataset. The final result is calculated by averaging the outputs of individual models built over each bootstrap set (7-9).
6) M5Rules.This algorithm generates a decision list for regression problems using separate-and-conquer. In each iteration, it builds a model tree and then generates a rule out of the "*best*" leaf. The algorithm divides the space of the parameters into subspaces and builds in each of them a linear regression model. It is based on the M5 algorithm. Once all the examples are covered, the algorithm terminates (7, 9).
7) M5P – Pruned Model Tree. The algorithm is based on decision trees; however, instead of having values at the tree's nodes, it contains a multivariate linear regression model at each node. The input space is divided into cells using training data and their outcomes; then, a regression model is constructed as a tree leaf in each cell (7, 10, 11).
8) Random forests (12) is a substantial modification of bagging that creates a large number of de-correlated trees and then averages them. On many problems, the performance of random forests is very similar to boosting, and they are simpler to train and tune.

References

1. Chai T. DRR. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development.

2. Smola AJ, Schölkopf B. A tutorial on support vector regression. Statistics and Computing. 2004; 14:199-222.

3. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK. Improvements to the SMO algorithm for SVM regression. IEEE Transactions on Neural Networks. 2000; 11:1188-93.

4. Ramyaa R, Hosseini O, Krishnan GP, Krishnan S. Phenotyping women based on dietary macronutrients, physical activity, and body weight using machine learning tools. Nutrients. 2019; 11.

5. Mitchell TM. 1997: McGraw-Hill Inc.: New York, NY, USA.

6. Wildeboer RR, van Sloun RJG, Wijkstra H, Mischi M. Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods. Comput Methods Programs Biomed. 2020; 189:105316.

7. Graczyk M, Lasota T, Trawiński B, Trawiński K, editors. Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. Intelligent Information and Database Systems; 2010 2010//; Berlin, Heidelberg: Springer Berlin Heidelberg.

8. Breiman L. Bagging predictors. Machine learning 1996; 24:123-140.

9. Holmes G, Hall M, Prank E. Generating rule sets from model trees. Advanced topics in artificial intelligence. In: Foo N. (eds) Advanced Topics in Artificial Intelligence. AI 1999. Lecture Notes in Computer Science, vol 1747. Springer, Berlin, Heidelberg.

10. Quinlan RJ. Learning with Continuous Classes. In: 5th Australian Joint Conference on Artificial Intelligence, Singapore. 1992; 343-348.

11. Wang Y, Witten IH. Induction of model trees for predicting continuous classes. Working Paper. 1996 Contract No.: 96/23.

12. Breiman L. Random Forests. Machine Learning. 2001; 45:5-32.