

Title: Learning under constraints with EXPLORE

Authors: Aniek F. Markus, Jan A. Kors, Peter R. Rijnbeek

Background

Many patient-level prediction models are developed and published in literature, but only few are used in clinical practice (1). EXPLORE (Exhaustive Procedure for LOGic-Rule Extraction) is an exhaustive search algorithm designed to find optimal decision rules (2). This algorithm has several features that make it attractive for patient-level prediction models. First, the resulting prediction model is a (short) decision rule and can thus be considered interpretable, which can contribute to create trustworthy AI (3). Second, the exhaustive search nature of the algorithm allows users to specify additional constraints on the model (e.g. restricting the rule length and forcing certain features to be included in the model) and/or performance (e.g. minimum specificity and sensitivity). We aim to investigate the potential of EXPLORE for patient-level prediction models developed in OHDSI by comparing the performance to two more frequently used methods for patient-level prediction.

Methods

EXPLORE (2) generates decision rules of pre-specified length in disjunctive normal form (DNF). A formula in DNF is a disjunction of terms (OR, \vee), where the terms are conjunctions (AND, \wedge) of literals, and the literals are feature-operator-value triples ($A > a$). An example of a DNF formula is $(A > a \wedge B = b) \vee C \geq c$, the resulting decision rule has the form: if (DNF formula) then class = 1 else class = 0.

To find the best decision rule EXPLORE performs an exhaustive search of all possible rules of pre-specified length using smart techniques to reduce the search space. For example, by reducing the number of values for each feature that need to be checked (subsumption pruning) and disregarding subspaces that cannot contain the optimal decision rule (branch-

and-bound). The exhaustive search approach guarantees we find an optimal decision rule and allows users to specify additional constraints while optimizing over a chosen performance metric. For more details on the EXPLORE algorithm we refer to the original publication (2).

We developed an R package to run EXPLORE, which is implemented in C++. We then investigated the performance of EXPLORE in comparison to two more frequently used methods for patient-level predictions: LASSO logistic regression and RandomForest. As default setting for EXPLORE we use a maximum rule length of 3 and maximize accuracy. In addition, we investigate learning with EXPLORE under two types of constraints:

- Minimum specificity 0.9,
- Mandatory feature, selecting the best predictor in LASSO logistic regression.

We test the performance of these methods on 13 standard datasets from the UCI data repository (4). Each dataset contains a class variable (the outcome) and 4 to 27 features. If the class variable is non-binary, we predict the majority class.

The set up of the evaluation is as follows. In each dataset we perform 3-fold cross validation, thus repeatedly training a prediction model in 2/3 of the data and evaluating the performance in the remaining 1/3. We measure prediction performance in terms of the area under the receiver operating characteristic curve (AUC). The performance is averaged over the different folds. In addition, we indicate the model size as the median number of non-zero features (LASSO logistic regression), number of features used in one of the trees (RandomForest), and rule length (EXPLORE).

Results

The developed R package to run EXPLORE can be downloaded from GitHub:

<https://github.com/mi-erasmusmc/explore/>. The performance in terms of AUC of the different

methods on standard UCI datasets is visualized in Figure 1. The figure shows that there is not one method systematically outperforming the others. Instead, the methods perform roughly similar on the same prediction problems, even though some prediction problems seem to be more difficult than others. For 3/13 prediction problems EXPLORE produces the best prediction model based on AUC.



Figure 1: Comparison performance of AUC on standard UCI datasets between LASSO logistic regression, RandomForest, and EXPLORE (maximum rule length 3).

When analyzing the model size based on the number of features used by the different methods, we observe larger differences. Figure 2 shows RandomForest uses most features (in fact all available features are often used), followed by LASSO logistic regression, and EXPLORE with the lowest number for 11/12 prediction problems.

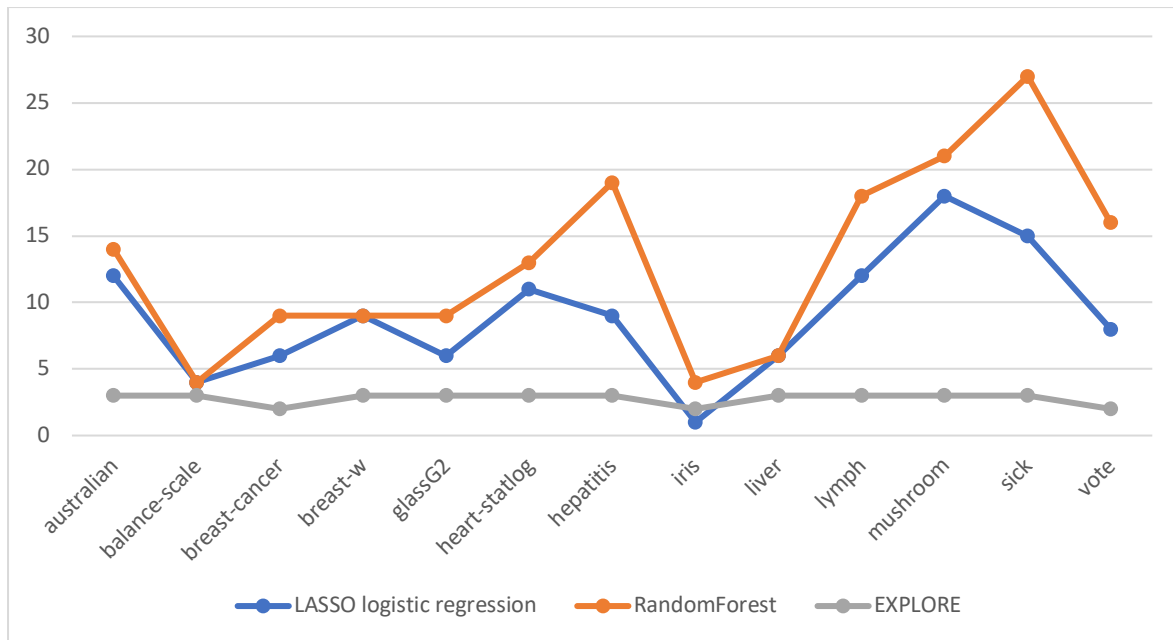


Figure 2: Comparison model size on standard UCI datasets between LASSO logistic regression, RandomForest, and EXPLORE (maximum rule length 3).

The results with EXPLORE under two types of constraints are also shown in Figure 1. Imposing a minimum specificity of 0.9 leads to a slightly lower AUC compared to the situation excluding this constraint, but allows the model developer to trade-off the sensitivity and specificity as desired. Adding one mandatory feature led to a different model in 4/13 prediction problems. In these cases, the AUC decreased only marginally (0.005-0.016) and increased once (0.01).

Conclusion

Results on standard UCI datasets show that EXPLORE can achieve similar performance for prediction problems compared to LASSO logistic regression and RandomForest, while the model size is substantially smaller. Moreover, we have demonstrated the ability of EXPLORE to learn under different types of constraints. Another benefit that should be explored in future research is the possibility to add mandatory features in the model based on domain knowledge to enhance the face value and generalizability of the resulting model.

The current results are limited in the sense that the studied prediction problems are simpler than real-world settings. Routinely collected health care data is typically more complex; it contains a much larger number of observations, thousands of features, and possibly more variation in the values of features. These are all factors that will influence the computational feasibility of EXPLORE. In the upcoming months we plan to investigate this further by adding EXPLORE to the PatientLevelPrediction package (<https://github.com/ohdsi/patientlevelprediction/>) and evaluating the feasibility and performance on real-world clinical prediction problems using the Integrated Primary Care Information (IPCI) database mapped to OMOP-CDM.

Funding

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806968. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA.

References

1. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med.* 2019;25(1):30-6.
2. Rijnbeek PR, Kors JA. Finding a short and accurate decision rule in disjunctive normal form by exhaustive search. *Machine learning.* 2010;80(1):33-62.
3. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *J Biomed Inform.* 2021;113:103655.
4. UCI Machine Learning Repository [Internet]. University of California, Irvine, School of Information and Computer Sciences. 2019. Available from: <http://archive.ics.uci.edu/ml>.