Development of a multi-model system to accommodate unknown misclassification costs in prediction of patient recruitment in multicentre clinical trials

Gilyana Borlikova, Michael O'Neill
Natural Computing Research & Applications Group
School of Business, University College Dublin
Dublin, Ireland
gilyana.borlikova@ucd.ie,m.oneill@ucd.ie

Louis Smith, Michael Phillips
ICON plc
Dublin, Ireland
louis.smith@iconplc.com,michael.phillips@iconplc.com

ABSTRACT

Clinical trials are an essential step in a new drug's approval process. Optimisation of patient recruitment is one of the major challenges facing pharma and contract research organisations (CRO) in conducting multicentre clinical trials. Improving the quality of selection of investigators/sites at the start of a trial can help to address this business problem. Grammatical Evolution (GE) was previously used to evolve classification models to predict the future patient enrolment performance of investigators/sites considered for a trial. However, the unknown target misclassification costs at the model development stage pose additional challenges. To address them we use a new composite fitness function to develop a multi-model system of decision-tree type classifiers that optimise a range of possible trade-offs between the correct classification and errors. The predictive power of the GE-evolved models is compared with a range of machine learning algorithms widely used for classification. The results of the study demonstrate that the GE-evolved multi-model system can help to circumvent uncertainty at the model development stage by providing a collection of customised models for rapid deployment in response to business needs of a clinical trial.

CCS CONCEPTS

•**Applied computing → Life and medical sciences;** *Health informatics;* Health care information systems;

ACM Reference format:
Gilyana Borlikova, Michael O'Neill and Louis Smith, Michael Phillips. 2017. Development of a multi-model system to accommodate unknown misclassification costs in prediction of patient recruitment in multicentre clinical trials . In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 2 pages.
DOI: http://dx.doi.org/10.1145/3067695.3076062

1 INTRODUCTION, PROBLEM DEFINITION AND BACKGROUND

Patient recruitment is the most time and resource consuming part of the majority of clinical trials [9]. In previous work we employed GE [7] to evolve classification model capable of predicting enrolment performance of clinical sites [3]. We have shown that GE is capable of evolving classifiers that are comparable or even better than results of the benchmark machine learning (ML) algorithms. However, standard accuracy fitness function does not take into account class distribution. In most real-life patient recruitment situations, classes are unbalanced and misclassification costs are different and neither characteristic is known at the model development stage. For such situations [8] proposed an idea of a robust hybrid classifier, the idea further developed by [1, 4, 6].

In this study we identify a range of potentially acceptable misclassification costs/thresholds for the negative class and use GE to evolve a system of models that accommodate these costs. To achieve this we develop a fitness function capable of accommodating to varying False Positive Rate (FPR) thresholds .

2 EXPERIMENTS, RESULTS AND ANALYSIS

The dataset was described previously in [3] and constructed based on the anonymised historical operational data provided by ICON plc (1233 records, 42 variables describing different characteristics of prospective investigator/site). The sites were allocated to two classes based on their patient recruitment performance. The data was split into train/test subsets (70/30%) and GE was used to evolve decision-tree type classifiers. The best of run GE models were tested to ascertain their generalisation ability. The GE grammar (similar to [3]) used the function and terminal set detailed in Table 1. The evolutionary parameters used were: population 1000 individuals, 50 generations, ramped-half-and-half initialisation, tournament selection (size 5), generational replacement, elite size 1, sub-tree crossover (90% probability), sub-tree mutation (1 event/individual), maximum tree depth 9, 30 independent runs.

We introduced a new fitness function to facilitate evolving solutions that maximise performance in terms of True Positive Rate (TPR) (True Positive/Condition Positive) given FPR (False Positive/Condition Negative) cut-off value:

$$Fitness = \begin{cases} TPR & \text{if } FPR \leq \text{cut-off} \\ -FPR & \text{if } FPR > \text{ cut-off} \end{cases}$$

Depending on the business environment, the site selection might benefit from either more conservative or more liberal models. We investigated four FPR cut-off values: 0.2, 0.3, 0.4, 0.5. The best evolved GE models were benchmarked against three well-established ML algorithms - Classification and Regression Tree (cart), Random Forest (rf) and Gaussian Support Vector Machines (svm). The

GECCO '17 Companion, Berlin, Germany
 978-1-4503-4939-0/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3067695.3076062

Table 1: Function and terminal sets of GE classifier

Function set	Terminal set
$+, -, *, /, \wedge, \vee, \neg$	35 numerical variables: x0, ..., x34
$=, \neq$	3 categorical variables: x35, x36, x37
$<, >, \leq, \geq$	4 Boolean variables: x38, ..., x42
	20 constants in -1.0, ..., 1.0 with 0.1 step

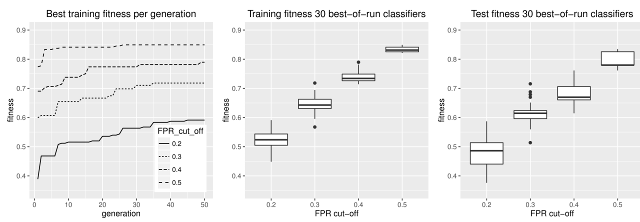

Figure 1: GE classification experiments. Best training fitness of GE models evolved with 4 FPR cut-offs (left), fitness of the 30 best-of-run GE classifiers on the training data (middle) and on the test data (right)

ML models were trained and tuned using 10 times 10-fold cross-validation using AUC and 0.5 class threshold (R CARET package, parameter settings: cart ($cp = 0$); rf ($\#predictors = 13$); svm ($\sigma = 0.0000129, cost = 512$)).

The best (Fig. 1 left) and average population fitness increased over 50 GE generations. The median training performance of the 30 best classifiers evolved with each FPR cut-off was 0.52, 0.64, 0.73, and 0.83 respectively (Fig. 1 middle). As expected, in comparison with the median performance on the train subset, the median TPR levels achieved by these models on the test subset were lower (0.49, 0.62, 0.67, 0.78 respectively), reflecting the challenge of generalisation (Fig. 1 right).

Training/test AUCs of the ML models were as follows: cart - 0.846/0.750, rf - 0.966/0.742, svm - 0.859/0.710. Class thresholds to satisfy FPR cut-offs based on the training data were selected and then applied to classification of the test data.

The results (Table 2) show that in all four experiments GE-evolved models maintain their positioning around FPR cut-off values more consistently than ML models with pre-selected class thresholds (in bold - actual FPR levels within ±0.05 of the desired FPR levels). Apart from the 0.2 FPR cut-off experiment, GE models achieve the highest TPR between models maintaining FPR positioning (in bold). Taken together, results of this study demonstrate that use of the new fitness function with different FPR cut-offs to drive GE generates models that uphold these FPR cut-offs on the test data.

2.1 Conclusions

This study approached the business problem of improving patient recruitment in multicenter clinical trials by developing predictive classification models of future performance of clinical sites. This problem is complicated by the unknown and/or changing misclassification costs. We used GE with a new fitness function that incorporates FPR threshold to evolve a system of classifiers that maximise

Table 2: Performance of models developed with different FPR cut-offs on test

Model	Metric	0.2	0.3	0.4	0.5 cut-off
ge	TPR	0.59	**0.72**	**0.75**	**0.83**
ge	FPR	**0.19**	**0.32**	**0.42**	**0.48**
cart	TPR	0.66	0.71	0.73	0.79
cart	FPR	**0.24**	**0.30**	**0.36**	0.44
rf	TPR	0.76	0.80	0.85	0.89
rf	FPR	0.32	0.43	0.54	0.64
svm	TPR	0.66	0.71	0.72	0.82
svm	FPR	0.33	0.38	0.39	0.55

correct identification of the class of interest while maintaining different levels of the other class misclassification. The resultant models show generalisation levels comparable with or even better than the well-established ML models, while maintaining the required levels of misclassification.

The same problem can be re-cast as a multi-objective optimisation problem. Several recent studies successfully used Evolutionary Multi-objective Optimisation (EMO) to solve similar problems [1, 2, 4–6]. We will investigate utility of an EMO approach in future work.

ACKNOWLEDGMENTS

This research is based upon work supported by ICON plc.

REFERENCES

[1] Simon Bernard, Clément Chatelain, Sébastien Adam, and Robert Sabourin. 2016. The Multiclass ROC Front Method for Cost-sensitive Classification. *Pattern Recogn.* 52, C (April 2016), 46–60. DOI: http://dx.doi.org/10.1016/j.patcog.2015.10.010

[2] Urvesh Bhowan, Mark Johnston, Mengjie Zhang, and Xin Yao. 2013. Evolving Diverse Ensembles Using Genetic Programming for Classification With Unbalanced Data. *Trans. Evol. Comp* 17, 3 (June 2013), 368–386. DOI: http://dx.doi.org/10.1109/TEVC.2012.2199119

[3] Gilyana Borlikova, Michael Phillips, Louis Smith, and Michael O'Neill. 2016. Evolving Classification Models for Prediction of Patient Recruitment in Multicentre Clinical Trials Using Grammatical Evolution. In *Applications of Evolutionary Computation. EvoApplications 2016. Lecture Notes in Computer Science, vol 9597*, Giovanni Squillero and Paolo Burelli (Eds.). Springer International Publishing, 46–57. DOI: http://dx.doi.org/10.1007/978-3-319-31204-0

[4] Clément Chatelain, Sébastien Adam, Yves Lecourtier, Laurent Heutte, and Thierry Paquet. 2010. A Multi-model Selection Framework for Unknown and/or Evolutive Misclassification Cost Problems. *Pattern Recogn.* 43, 3 (March 2010), 815–823. DOI: http://dx.doi.org/10.1016/j.patcog.2009.07.006

[5] Rui Dilão, Daniele Muraro, Miguel Nicolau, and Marc Schoenauer. 2009. Validation of a Morphogenesis Model of Drosophila Early Development by a Multi-objective Evolutionary Optimization Algorithm. In *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics. EvoBIO 2009. Lecture Notes in Computer Science, vol 5483*, Clara Pizzuti, Marylyn D. Ritchie, and Mario Giacobini (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 176–190.

[6] Julien-Charles Levesque, Audrey Durand, Christian Gagne, and Robert Sabourin. 2012. Multi-objective Evolutionary Optimization for Generating Ensembles of Classifiers in the ROC Space. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (GECCO '12)*. ACM, New York, NY, USA, 879–886.

[7] Michael O'Neill and Conor Ryan. 2003. *Grammatical Evolution: Evolutionary Automatic Programming in a Arbitrary Language.* Genetic programming, Vol. 4. Kluwer Academic Publishers.

[8] Foster Provost and Tom Fawcett. 2001. Robust Classification for Imprecise Environments. *Machine Learning* 42, 3 (2001), 203–231. DOI: http://dx.doi.org/10.1023/A:1007601015854

[9] Peter Schuler and Brendan Buckley. 2014. *Re-Engineering Clinical Trials: Best Practices for Streamlining the Development Process.* Academic Press.