Comparing Multi-class Classifiers: On the Similarity of Confusion Matrices for Predictive Toxicology Applications

Mokhairi Makhtar, Daniel C. Neagu, and Mick J. Ridley

School of Computing, Informatics and Media, University of Bradford, Bradford, BD7 1DP, UK
{M.B.Makhtar,D.Neagu,M.J.Ridley}@Bradford.ac.uk

Abstract. Calculating the similarity of predictive models helps to characterize the models diversity and to identify relevant models from a collection of models. The relevant models are considered based on their performance, calculated using their confusion matrix. In this paper, we propose a methodology to measure the similarity for predictive models performances by comparing their confusion matrices. In this research, we focus on multi-class classifiers for toxicology applications. The performance measures of confusion matrices of multi-class classifiers are regrouped into a binary classification problem. Such approach may result in selecting multi-class classifiers with lower False Negative Rate (FNR) for example. Consequently, the methodology for model comparison based on the similarity of confusion matrices provides a working way to select models from a collection of classifiers.

Keywords: Similarity of Confusion Matrices, Classifiers Comparison, Multi-Class Classifiers.

1 Introduction

Predictive models comparison helps in finding how similar models are. But relying only on standard performance indicators such as accuracy may not give much clue on the overall or specific quality of a predictive model. Sometimes the accuracy might be biased for a certain class and this may not provide a good indication of the overall performance for the predictive model. In this case the accuracy is not necessarily the best measurement for predictive models, whereas the confusion matrix is still the most valuable source of performance indicators from classifiers to be analyzed.

Our motivation is given by the need of analyzing the multi-class classifier models for selected classes. In toxicology, we are mostly interested in the toxic class being predicted correctly. Using the confusion matrix as the information source of classifiers performance, we can adapt more useful measurements related to our objective. The classifiers can be either binary class or multi-class models. In our case, we want to predict if the chemical compound is toxic or non-toxic where all our

H. Yin, W. Wang, and V. Rayward-Smith (Eds.): IDEAL 2011, LNCS 6936, pp. 252–261, 2011.

classifiers are in a multi-class format. The multi-class classifiers can be used as binary class models. It is done by combining the multi-class dataset into a new dataset with only binary classes of toxic and non-toxic output [1, 2] and re-generate new predictive models related to the new datasets. But the solution requires much effort in converting datasets to new binary class sets and retraining the models with the new datasets. To be more practical because there are thousands of models in a collection of models, we propose to use the multi-class classifiers confusion matrices as new binary class classifiers confusion matrices. The practical method is to transform the multi-class confusion matrices into binary confusion matrices without updating the datasets and re-generating the models. This will confirm that the original structures and information the predictive models learned remain unchanged. We will demonstrate the proposed technique in section 3.

In this paper we propose a technique to compare multi-class predictive models' performance measures based on confusion matrices. Our methodology addresses model selection, where comparing the classifiers' performance for each class will lead to usefully diverse predictive models for the class of interest from model ensembles.

The rest of the paper is structured as follows: Section 2 presents related work on reducing multi-class into binary class problem. Section 3 defines the technique proposed for comparison of confusion matrices for multi-class (toxicology) models. In Section 4 we introduce and exemplify the technique to calculate the performance measures of output for multi-class predictive models represented by their confusion matrix. Experiments and results are discussed in Section 5. The paper ends with conclusions on current work and further research directions.

2 Reducing Multi-class to Binary Classification Problems

Sometimes, the multi-class classification problems can still be solved with binary classifiers. Such a solution may divide the original multi-class dataset into two class subsets, learning a different binary model for each subset. These techniques are known as binarisation strategies. There are three main approaches: *One-vs-All* (OVA), *One vs-One* (OVO), and *Error Correcting Output Codes* (ECOC) [2].

All of these techniques decompose a complex multi-class to a simpler binary class problem. Hence this strategy may improve the performance because the classifiers have an easier task to distinguish between only two classes rather than many classes.

In this paper we want to investigate whether there are any differences in performance between binarisation strategies by regenerating new binary classifiers from multi-class classifiers. We calculate the performance measures using multi-class classifiers confusion matrices without retraining new binary classifiers.

In the next section, we will discuss on the performance measures related to binary classification classifiers and propose a methodology to reduce multi-class problems to a binary version while calculating the performance measures of the multi-class classifiers with a focus on lower False Negative Rate (FNR) for example, as required in toxicity prediction problems.