

A Hybrid Evaluation Metric for Optimizing Classifier

Hossin M.^{*,#,1}, Sulaiman M.N.^{*,2}, Mustapha A.^{*,3}, Mustapha N.^{*,4}, Rahmat R.W.^{*,5}

^{*}Faculty of Computer Science and Information Technology
Universiti Putra Malaysia (UPM)
43400, UPM Serdang, Selangor, Malaysia

[#]Faculty of Cognitive Sciences and Human Development
Universiti Malaysia Sarawak (UNIMAS)
93400 Kota Samarahan, Sarawak, Malaysia

¹mhossin78@gmail.com, {²nasir, ³aida, ⁴norwati, ⁵rahmita}@fsktm.upm.edu

Abstract - The accuracy metric has been widely used for discriminating and selecting an optimal solution in constructing an optimized classifier. However, the use of accuracy metric leads the searching process to the sub-optimal solutions due to its limited capability of discriminating values. In this study, we propose a hybrid evaluation metric, which combines the accuracy metric with the precision and recall metrics. We call this new performance metric as Optimized Accuracy with Recall-Precision (OARP). This paper demonstrates that the OARP metric is more discriminating than the accuracy metric using two counter-examples. To verify this advantage, we conduct an empirical verification using a statistical discriminative analysis to prove that the OARP is statistically more discriminating than the accuracy metric. We also empirically demonstrate that a naive stochastic classification algorithm trained with the OARP metric is able to obtain better predictive results than the one trained with the conventional accuracy metric. The experiments have proved that the OARP metric is a better evaluator and optimizer in the constructing of optimized classifier.

Keywords - Hybrid evaluation metric, Accuracy metric, Precision metric, Recall metric, Classifier optimization

I. INTRODUCTION

The definition of classification can be varied. In a simple definition, classification can be defined as a process of assigning data or instances to a particular pre-specified fixed number of classes according to its predicted class by measuring a number of attributes on a particular data or instances. Technically, classification process involves two main phases. Firstly, classification process starts with building an optimized classifier through a training data set. Secondly, the optimized classifier is used for testing, which is used to predict the unknown data. In this study, we are interested to investigate the use of new evaluation metric in building the optimized classifier during the training phase.

To produce an optimized classifier, certain metrics are used to evaluate the generated solutions. The most common and popular evaluation metric used for evaluating and selecting the optimal solution is accuracy or error rate (1-

accuracy). However, using accuracy metric as benchmark measurement has limitations.

Let us consider two solutions $s=\{A, B\}$ in a two-class confusion matrix where both A and B produce an equivalent total of correct predicted instances as shown in Table I. However, as indicated in Table I, both solutions exhibit the same accuracy, which is 95%. In this case, the accuracy metric could not discriminate whether A or B is better. This clearly indicates that the accuracy metric exhibit poor discriminating value to discriminate the optimal solution due to less information about the overall results. Nonetheless, intuitively, we can conclude that solution B is better than A . This is because the total instances that have been correctly predicted by A for both classes are approximately balanced (48:47). In contrast, the total of instances that correctly predicted by B is unbalanced (50:45) for both classes.

TABLE I THE PROBLEM WITH ACCURACY FOR BALANCED CLASS DISTRIBUTION (50:50)

Actual Class Predicted class	Solution A		Solution B	
	Pos	Neg	Pos	Neg
Pos	50	5	48	3
Neg	0	45	2	47

On top of that, few studies have reported that the simplicity of the accuracy metric can lead the selection and the discrimination processes to the sub-optimal solutions especially when dealing with imbalanced class instances [1][2][3][4]. This is because a small class of instances has very little impact on the accuracy as compared to the large class of instances. Clearly, this indicates that the performance of accuracy metric is not robust and can be drastically affected by the changes in data proportion.

On the other hand, the precision and recall are two evaluation metrics that are commonly used as the alternative metric to measure the performance of binary classifiers for two different aspects [5]. Basically, precision is used to determine the fraction of positive instances that are correctly predicted in a positive class, while recall measures the