Journal of Computer Science 7 (4): 582-590, 2011
ISSN 1549-3636


# A Novel Performance Metric for Building an Optimized Classifier

[1,2]Mohammad Hossin, [1]Md Nasir Sulaiman,
[1]Aida Mustapha and [1]Norwati Mustapha
[1]Department of Computer Science,
Faculty of Computer Science and Information Technology,
University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia
[2]Department of Cognitive Science,
Faculty of Cognitive Sciences and Human Development,
University Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia

**Abstract: Problem statement:** Typically, the accuracy metric is often applied for optimizing the heuristic or stochastic classification models. However, the use of accuracy metric might lead the searching process to the sub-optimal solutions due to its less discriminating values and it is also not robust to the changes of class distribution. **Approach:** To solve these detrimental effects, we propose a novel performance metric which combines the beneficial properties of accuracy metric with the extended recall and precision metrics. We call this new performance metric as Optimized Accuracy with Recall-Precision (OARP). **Results:** In this study, we demonstrate that the OARP metric is theoretically better than the accuracy metric using four generated examples. We also demonstrate empirically that a naïve stochastic classification algorithm, which is Monte Carlo Sampling (MCS) algorithm trained with the OARP metric, is able to obtain better predictive results than the one trained with the conventional accuracy metric. Additionally, the t-test analysis also shows a clear advantage of the MCS model trained with the OARP metric over the accuracy metric alone for all binary data sets. **Conclusion:** The experiments have proved that the OARP metric leads stochastic classifiers such as the MCS towards a better training model, which in turn will improve the predictive results of any heuristic or stochastic classification models.

**Key words:** Performance metric, Optimized Accuracy with Recall-Precision (OARP), accuracy metric, extended precision, extended recall, optimized classifier, Monte Carlo Sampling (MCS)

## INTRODUCTION

To date, many efforts have been carried out to design more advanced algorithms to solve classification problems. At the same time, the development of appropriate performance metrics to evaluate the classification performance are at least as importance as algorithm. In fact, it is a key point to produce a successful classification model. In other words, the performance metric plays a significant role in guiding the design of better classifier.

From the previous studies, the performance metric is normally employed in two stages (i.e., the training stage and the testing stage). The use of performance metric during the training stage is to optimize the classifier (Ferri *et al*., 2002; Ranawana and Palade, 2006). In other words, in this particular stage, the performance metric is used to discriminate and to select the optimal solution which can produce a more accurate prediction of future performance. Meanwhile, in the testing stage, the performance metric is usually employed for comparing and evaluating the classification models (Bradley, 1997; Caruana and Niculescu-Mizil, 2004; Kononenko and Bratko, 1991; Provost and Domingos, 2003; Seliya *et al*., 2009).

In this study, we are interested about the use of performance metric in evaluating and building an optimized classifier for any heuristic and stochastic classification algorithms. In general, these algorithms use the training stage learns from the data and at the same time attempt to optimize the solution by discriminating the optimal solution from the large space of solutions. In order to find the optimal solution, the

**Corresponding Author:** Mohammad Hossin, Department of Computer Science, Faculty of Computer Science and Information Technology, University Putra Malaysia, 43400 Serdang, Selangor, Malaysia