A REVIEW ON EVALUATION METRICS FOR DATA CLASSIFICATION EVALUATIONS

Hossin, M.[1] and Sulaiman, M.N.[2]

[1]Faculty of Computer Science & Information Technology, Universiti Malaysia Sarawak,
94300 Kota Samarahan, Sarawak, Malaysia
[2]Faculty of Computer Science & Information Technology, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

Evaluation metric plays a critical role in achieving the optimal classifier during the classification training. Thus, a selection of suitable evaluation metric is an important key for discriminating and obtaining the optimal classifier. This paper systematically reviewed the related evaluation metrics that are specifically designed as a discriminator for optimizing generative classifier. Generally, many generative classifiers employ accuracy as a measure to discriminate the optimal solution during the classification training. However, the accuracy has several weaknesses which are less distinctiveness, less discriminability, less informativeness and bias to majority class data. This paper also briefly discusses other metrics that are specifically designed for discriminating the optimal solution. The shortcomings of these alternative metrics are also discussed. Finally, this paper suggests five important aspects that must be taken into consideration in constructing a new discriminator metric.

KEYWORDS

Evaluation Metric, Accuracy, Optimized Classifier, Data Classification Evaluation

1. INTRODUCTION

In data classification problems, data can be divided into commercial data, texts, DNAs and images. This paper emphasizes on commercial data as the focus of discussion. Furthermore, data classification can be divided into binary, multiclass and multi-labelled classification [33]. In this paper, the study is aimed on binary and multiclass classification which focuses on the evaluation metrics for evaluating the effectiveness of classifiers. In general, the evaluation metric can be described as the measurement tool that measures the performance of classifier. Different metrics evaluate different characteristics of the classifier induced by the classification algorithm.

From the literature, the evaluation metric can be categorized into three types, which are threshold, probability and ranking metric [2]. Each of these types of metrics evaluates the classifier with different aims. Furthermore, all these types of metrics are scalar group method where the entire performance is presented using a single score value. Thus, it makes easier to do the comparison and analysis although it could mask subtle details of their behaviours. In practice, the threshold and ranking metric were the most common metrics used by researchers to measure the performance of classifiers. In most cases, these types of metrics can be employed into three different evaluation applications [23].

First, the evaluation metrics were used to evaluate the generalization ability of the trained classifier. In this case, the evaluation metric is used to measure and summarize the quality of trained classifier when tested with the unseen data. Accuracy or error rate is one of the most