Loss functions for optimizing Kappa as the evaluation measure for classifying diabetic retinopathy and prostate cancer images

Rajendran Nirthika Siyamalan Manivannan Amirthalingam Ramanan
Department of Computer Science, Faculty of Science, University of Jaffna, Jaffna, Sri Lanka.
nirthika.rajendran@gmail.com siyam@univ.jfn.ac.lk a.ramanan@univ.jfn.ac.lk

***Abstract*—Quadratic Weighted Kappa (QWK) is a statistic to measure the agreement between two annotators. QWK has been widely used as the evaluation measure for various medical imaging problems, where, the class labels have a natural ordering, e.g., no Diabetic Retinopathy (DR), mild DR, and severe DR. The easiest way to treat the classification problem with these ordinal labels is to consider the problem as a multiclass classification problem and apply the Cross Entropy (CE) loss. However, when applying CE loss the order of the classes becomes meaningless, i.e., the loss will be same if a healthy image is classified into mild DR or severe DR. At the same time, the QWK score will be severely affected if a healthy image is classified into severe DR than mild DR. The most appropriate way to get a better classification score is to optimize the evaluation measure itself. i.e., directly optimize the QWK statistics. However, this optimization may hinder the learning, and may lead to sub-optimal solutions, and therefore, may give lower performance than expected. On the other hand, Ordinal Regression (OR) based approaches also can be used for such problems. The main focus of this work is to investigate which loss function (CE loss, QWK loss or OR loss) is the most appropriate one to the Convolutional Neural Network-based ordinal classification problems, where, QWK is used as the evaluation measure. Experiments on two public datasets, Diabetic Retinopathy and Prostate Cancer, with two different network architectures suggest that directly optimizing QWK is the better choice when small networks are used. On the other hand, we found that for large networks OR based loss function gives better performance.**

***Index Terms*—Convolutional neural network, Diabetic retinopathy, Image classification, Loss functions, Prostate cancer.**

I. Introduction

Convolutional Neural Networks (CNN) became the state-of-the-art methods for various medical imaging problems, including diabetic retinopathy [1]–[3] and cancer image classification [4], [5], due to their ability to jointly learn discriminative features and classifier from large image datasets. Usually the parameters of the network are learned using Stochastic Gradient Descent by back-propagation based on a loss function, which measures the degree of misclassification made by the network and penalize the wrongly classified instances. Although there are vaious loss functions used in the literature of deep learning for different problems, they differ from each other in various aspects, including, the considered problem and its difficulty, the way they penalize the misclassified instances, their robustness in learning against labeling noise, etc.

The *Cross Entropy* (CE) loss function is the widely used loss function used with deep neural networks for multiclass classification problems, where, usually the *Classification Accuracy* is used as the evaluation measure. This measure merely indicates how many instances are correctly classified, and how many are not. In the discrete multi-class classification problems the labels do not indicate any explicit ordering, e.g., classifying cats vs dogs. However, in some medical imaging domain the ordering of the labels is crucial, e.g., classifying images into different grades of abnormality. There is a natural ordering in the grades, for example, healthy images have no Diabetic Retinopathy (DR), mild DR is the early stage of the disease in which symptoms will be mild or nonexistent, and Proliferate DR is more advanced form of the disease. As the treatment is based on the grade of the abnormality, the labels of the images should be correctly predicted. For these ordinal classification problems neither CE loss as the loss function nor classification accuracy as the evaluation measure are the ideal choice as they entirely discard the ordering in the labels. On the other hand, *Quadratic Weighted Kappa* (QWK) is used to measure the agreement between annotators (or system vs. annotator) for ordinal classification problems in medical imaging domain, e.g., for classifying DR images [6], and classifying Prostate Cancer (PC) images [7]. *Ordinal Regression* (OR) based approaches which tries to minimize the distances between the actual labels and their predicted values are used in the literature for both DR and PC image classification. However, one could think that directly optimizing the evaluation measure, i.e., QWK, is the best choice to achieve the better QWK score. To best of our knowledge, there is a very little attention [8]–[10] was made to directly optimize QWK as the loss function.

Most of the existing work on DR [1]–[3], [11]–[14] and PC [5], [15]–[17] mainly focus on proposing/applying different network architectures for getting better performance [3], [11], [12], [14]–[16] or identifying the discriminative regions [1], [2], [5], [13]. Mostly, either CE loss [14], [17] or OR loss [1]–[3], [5], [11] was used for these problems. Differently from these, [8] and [9] try to directly optimize QWK as the loss function. For example, [8] shows slight improvements by QWK loss than optimizing the network over OR loss, and [9] reports significant improvement by QWK loss over the CE loss.