Pattern Recognition 44 (2011) 350–362

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

The ROC manifold for classification systems ☆

Christine M. Schubert [a,*], Steven N. Thorsen [b], Mark E. Oxley [a]

[a] Department of Mathematics and Statistics, Graduate School of Engineering and Management, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433-7765, United States
[b] Department of Mathematical Sciences, United States Air Force Academy, 2354 Fairchild Drive, Colorado Springs, CO 80840, United States

ARTICLE INFO

Article history:
Received 21 November 2008
Received in revised form
28 May 2010
Accepted 29 July 2010

Keywords:
Classification
Multiple classes
Receiver operating characteristic (ROC) curve
ROC manifold
Bayes cost

ABSTRACT

We define the ROC manifold and CC manifold as duals in a given sense. Their analysis is required to describe the classification system. We propose a mathematical definition based on vector space methods to describe both. The ROC manifolds for n-class classification systems fully describe each system in terms of its misclassifications and, by conjunction, its correct classifications. Optimal points which minimize misclassifications can be identified even when costs and prior probabilities differ. These manifolds can be used to determine the usefulness of a classification system based on a given performance criterion. Many performance functionals (such as summary statistics) preferred for CC manifolds can also be evaluated using the ROC manifold (under certain constraints). Examples using the ROC manifold and performance functionals to compete classification systems are demonstrated with simulated and applied disease detection data.

Published by Elsevier Ltd.

1. Introduction

Paramount to the development of classification systems is the ability to judge the usefulness of the system, whether judging the system against a benchmark level of acceptable performance or comparing it to other candidate systems. To make this judgement a performance criterion is required. One of the oldest and most commonly used performance tools used in the analysis of classification systems is the receiver operator characteristic (ROC) curve. The most commonly used ROC curve depicts the trade-off in correct classification for one pivotal class with the false classification into that class. A less common ROC curve depicts the trade-off of the two types of false classifications that can occur.

In the last decade, complexity in classification applications has warranted an extension of ROC curves and their analyses to describe and analyze systems in which there are three or more classes [1–11,14]. These extensions of ROC curves have produced various surfaces defined in terms of the correct classifications with the notable exception of [8,9,14], in which surfaces related to the misclassification errors are described. Points lying on these surfaces correspond to different operating parameters associated with the classification system. Often these parameters are thresholds (one example would be signal-to-noise ratio), though they need not be. There is no standardization of these surfaces and most focus on permutations of the correct classifications. For classification systems with three classes, these surfaces may be visualized in a three-dimensional plot of the true (correct) classification rates [1–7]. Since these surfaces are topological manifolds, we refer to them as correct classification manifolds (CC manifolds). For $n>2$ classes, concepts related to these surfaces have been proposed, many still focusing on the correct classification rates, though the increased dimensionality makes it impossible to view all correct classifications simultaneously [10]. At best for the n-class system, sets of three-dimensional plots can be used to examine the correct classifications for three classes at a time.

Initially, focusing on correct classification rates seems appealing since, for the three-class classification system, the trade-off between correct classifications can be compared graphically using each class's correct classification as an axis. Furthermore, summary measures of these CC manifolds focus on how well the classification systems correctly classify into their class states, thereby describing the overall correct classification rate. By conjunction, then, the overall misclassification rate for the entire system is described, although no information is directly obtainable about misclassifications within each class. Such summary measures include the total correct classification rate and volume under the surface (VUS) [1,6,9]. Many researchers have examined VUS for systems with more than two classes [12,3,10,13,6,14,4] with the view of constructing a polytope from the data to

☆ The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the US Government.

* Corresponding author.
E-mail addresses: christine.schubertkabban@afit.edu (C.M. Schubert), Steven.Thorsen@usafa.edu (S.N. Thorsen), Mark.Oxley@afit.edu (M.E. Oxley).

0031-3203/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.patcog.2010.07.025

calculate or describe how to use the VUS. The appeal of VUS is that this summary performance quantifier hopefully becomes a probability estimate as it does with the two-class case, generalizing the diagnostic ability of the classification system across all operating parameters. For a CC manifold, this can be interpreted as the chance of correct classification when presented with, as a group, one randomly selected subject from each class [1,6]. To illustrate further using medical diagnostics, the resulting VUS for three classes may be interpreted as the probability that a clinician diagnoses each individual to the correct diagnostic class after being presented with three individuals from three different classes. Herein lies more confusion, however, because in the VUS defined by the volume under a polytope created in the space based on classification errors, the probability of correct classification is not necessarily 1-VUS, if VUS indeed exists. In contrast, for the VUS defined by the volume under a polytope created in the space based on correct classifications, the probability of correct classification is the VUS.

There are noted issues surrounding the use of VUS [15,9]. In [15] we see that conclusions made when comparing classification systems based on VUS infer the classification system's diagnostic ability, with the caveat that these calculations assume equal weighting for prior probabilities and costs between the classes. However, there are no costs associated with correct classifications, only errors in classifications, and as such, summary statistics not considering misclassifications cannot address these costs. In [8,9] we see a definition of a ROC hypersurface and the hypervolume under it which extended previous efforts beyond the three-class case to an n-class case. It is demonstrated that the "guessing" (and through convergence the "near-guessing") observer has the same VUS as the "perfect" ideal observer.

As a result of these works, there are two important issues to address. First, there is a dual problem in the CC manifold. Given an n-class classification system, analysis of the dual problem involves an $(n-1)$-dimensional linear variety of the n-space containing the CC manifold. Since this linear variety is codimension 1 to the correct classification space (CC space), a surface can always be generated under it (ignoring the second issue discussed below). Therefore, [9] would have the ROC hypersurface VUS of every "perfect" ideal and "guessing" observer equal to 0. However, the CC manifold VUS of the "perfect" ideal observer is 1 in every case. This occurs because the surface created by the "guessing" observer will always be an n-simplex for this observer. For example, in the simpler two-class system which produces a ROC curve, we have $n=2$. Hence, the ROC space is in a space of dimension $2(2-1)=2$ while the ROC curve is isomorphic to a subset of the space $\mathbb{R}^1$, a space of dimension $2-1=1$, making the curve codimension 1 to the original ROC space. This creates a "volume" under the ROC curve. Notice also that the CC curve for the two-class system is also isomorphic to $\mathbb{R}^1$, which is codimension 1 to CC space. Thus, it too has a "volume". Of course, in these dimensions the "volume" is really area under the curve. This phenomenon is unique to the two-class case. Extending to a three-class case, the ROC space is a hypercube subset of $\mathbb{R}^{3^2-3}=\mathbb{R}^6$, while the CC space is a hypercube inside $\mathbb{R}^3$. The "guessing" observer is a classifier which is a subset of $\mathbb{R}^{3-1}=\mathbb{R}^2$ in ROC space. This clearly has no volume since the linear variety has codimension 4 to the ROC space; however, the guessing observer yields a 3-simplex in CC space, which has a volume of $\frac{1}{3!}=\frac{1}{6}$. These examples can be extend to any n-class system to demonstrate the existence of codimensions >1 which will suffer with similar problems. Further, these examples assume much in the dimensionality and independence of the underlying parameter spaces. Under ideal circumstances where there exist five independent parameters of the classification system, which vary as five of the six conditional probabilities of misclassification, the ROC manifold will be isomorphic to a linear variety in $\mathbb{R}^{3^2-3-1}=\mathbb{R}^5$, which is codimension 1 to ROC space. The second issue to address involves the importance of the parameters a classification system uses. In a three-class example, suppose we have less than five parameters (an occurrence that is acknowledged in[8,9]). Then the codimensionality of the space associated with the ROC manifold will be higher than 1, and no surface can exist. This is a very real possibility! In fact, the dimensionality of the problem has more to do with the underlying parameters of the classification system than with the number of classes, or independent misclassifications.

In this paper, we define the ROC manifold and CC manifold as duals in a given sense. Their analysis is required to describe the classification system. We propose a mathematical definition based on vector space methods to describe both. Unlike previous works, this definition makes no assumption that underlying distributions are known and thus can be utilized when likelihood decision criterion is unavailable. The ROC manifold for n-class classification systems fully describes the system in terms of its misclassifications and, by conjunction, its correct classifications. These manifolds can be used to determine the usefulness of a classification system based on a given performance criterion. We offer the ROC manifold not as a means for finding the optimal classifier through the use of utility or other criteria, but as a means to describe the performance of specific classification systems and to eventually compare performance between systems. Some performance functionals (such as summary statistics) useful for CC manifolds can also be evaluated using the ROC manifold (under certain constraints). Further, the ROC manifold may be computed regardless of the codimension that results from the possible classification systems, that is, directly, without the need to reduce parameters or dimensionality to create a manifold that is codimension 1 to the ROC space. Therefore, the definition of the ROC manifold may subsume previous ROC surface definitions in many cases. Another key difference of the ROC manifold with respect to CC surfaces is that optimal operating parameters may be identified when prior probabilities or costs differ among the various classes. In this paper, we will use the term, *parameter*, to refer to those continuous deterministic quantities that represent different settings for the classification system. These parameters are varied to compare system performance constituting the various points of the ROC manifold. The ROC manifold and CC manifold are paramount to fully evaluating the performances of the classification systems, and herein we endeavor to define them mathematically and describe them in detail.

This paper is constructed as follows. Section 2 outlines the necessary classification system theory. Section 3 defines the ROC manifold and the CC manifold. In this section we observe the relationship between the ROC manifold and the typical ROC curve when only two classes are of interest and between previous surfaces focusing only on correct classifications. We assume underlying distributions are not known and, therefore, likelihood decision criterion is unavailable. We also assume ROCs are invariant with respect to the prevalences of the various classes to be distinguished among, so that the class-conditional probabilities do not change if, or when, prior probabilities change. Section 4 details performance functionals useful for competing two or more classification systems and, specifically, focuses on Bayes cost as a decision criterion. In Section 5, we demonstrate the ROC manifold as useful in finding points of optimal performance defined in terms of the associated misclassification costs and prior probabilities. Using a simple classification system, Section 5 also gives examples that demonstrate the calculation of the ROC manifold and associated optimal points for codimension 1 and higher systems as well as illustrate some properties of this