Using ROC Curve In the Absence of Positive Examples

Zehra Cataltepe, Ming-Wei Chang, Thomas Stiefmeier, Bo-Juen Chen and Claus Neubauer
Siemens Corporate Research Inc.
755 College Road East
Princeton, NJ 08540
{zehra, mchang, thomas, bchen, neubauer}@scr.siemens.com

Abstract— **Receiver Operating Characteristics (ROC) curves have long been used to evaluate classifier performance in many fields (e.g. signal detection and machine learning). The ROC curve provides information on the tradeoff between the hit rate (true positives) and the false alarm rates (false positives). In order to draw the ROC curve both positive and negative examples are needed.**

In some applications, for example, machine condition monitoring, cancer detection, there are plenty of negative examples. However the positive examples are either rare, or do not fully describe the overall set of the possible positive examples. However, instead of the positive examples, some rules about the positive examples are available. For example, in machine condition monitoring, if a sensor drifts off from the set of observed states by a certain amount, we know that a fault has occurred.

In order to use the ROC curve to evaluate classifiers, we artificially create the positive examples based on the application dependent rules and the existing negative examples. Then we draw the ROC curve using this set of positive and negative examples.

I. Introduction

Receiver Operating Characteristics (ROC) curves have long been used to evaluate classifier performance in many fields [2, 5, 9] The ROC curve provides information on the tradeoff between the hit rate (true positives) and the false alarm rates (recalls or false positives).

After a classifier is trained using the training set, it is tested on a test set. Let the test inputs be $x_i \in R^k$ and the test outputs be $t_i, \in \{\text{True}, \text{False}\}, i = 1 \ldots, M$. Let the classifier (model) outputs for the ith test input be and $y_i \in \{\text{True}, \text{False}\}$. Then the true positive (TP) rate (or sensitivity) is defined as:

$$TP\ rate = \frac{\sum_{i=1}^{M} t_i = True \text{ AND } y_i = True}{\sum_{i=1}^{M} t_i = True}$$

The false positive (FP) rate (or 1-specificity) is:

$$FP\ rate = \frac{\sum_{i=1}^{M} t_i = False \text{ AND } y_i = True}{\sum_{i=1}^{M} t_i = False}$$

The ROC curve is produced by showing the FP rate on the x axis and the TP rate on the y axis. A classifier with a single TP and FP rate corresponds to a point on the ROC curve.

Most classifiers have internal threshold parameters that affect whether the output is True or False. By moving the threshold parameter an ROC curve is obtained for a classifier for different threshold values.

Area under the ROC curve (AUC) (for example [1, 4]) is the sum of the ROC values for a classifier when the threshold value changes between certain two values. Area under the ROC is between 0.5 (random guessing) and 1.0 (perfect classifier). AUC can be used to compare performance of different classifiers.

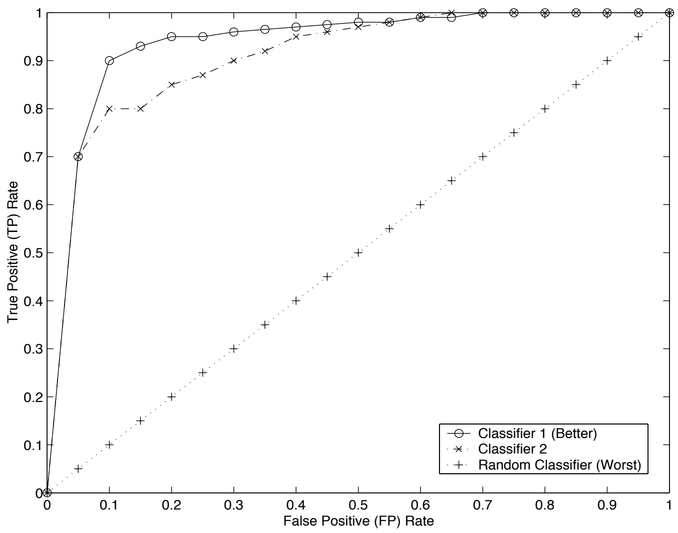

Fig. 1. Classifier 1 is better than 2 and 2 is better than 3. Classifier 3 has an AUC of 0.5.

Recently ROC area has been investigated by different authors. In [7] Provost et.al. showed that for comparison of different classifiers, ROC is better than classification accuracy. In [8] Provost et.al. used the ROC convex hull (ROCCH) method to compare classifiers under different cost and class distributions. They also investigated methods of combining classifiers to end up with a robust hybrid classifier. In [6] Mozer et.al. worked on designing a classifier that achieves a certain TP and FP rate based on the domain requirements. In [3] Fawcett investigated different strategies for evaluating rule sets when the goal is to maximize the ROC performance.

The rest of the paper is organized as follows: In section 2 we describe the machine condition monitoring problem and describe how to extend it so that the ROC analysis can be used to compare different models. In section 3 we describe our algorithm for producing positive examples based on a clean (all negative) data set and user determined rules (thresholds per sensor). In section 4 we show the ROC curves computed this way for a two sensor machine condition monitoring problem for different nearest neighbor classifiers. We conclude the paper with discussions in section 5.

II. Machine Condition Monitoring

In machine condition monitoring, the machine are operated under normal conditions and sensor data are collected. This sensor data is used for training and calibrating (testing) a model that best describes the data set. Any future sensor reading significantly different from the training/calibration set needs to be caught up as early as possible. (Please see figure 2.)

Although the goal is to be notified if the sensor readings ($\in R^k$) are not normal, it is also very important to know which sensor reading is not normal. Hence, instead of a $R^k \rightarrow R$ mapping, the model needs to do an $R^k \rightarrow R^k$ (auto-associative) mapping, indicating which sensor went wrong. Hence, the ROC analysis for a model consists of k ROC curves, one curve per sensor. When comparing different models, a weighted sum of the cost of these ROC curves can be used. We define cost at a certain TP, FP value as $Cost(FP, TP) = C_n * (1 - TP) + C_p * FP$ where C_n and C_p are the costs of false negatives and false positives respectively. In machine condition monitoring we are interested not on the whole ROC area, but only area around the TP values that correspond to the thresholds given by the analyst.

Machine condition monitoring data is very unbalanced, in the sense that the positive (faulty sensor) data is actually not in the training data. The analyst tries to assess the performance of a model by adding disturbances to the training data and noting if the model can recognize the added disturbance early enough. Early enough is usually expressed in terms of a threshold per sensor.

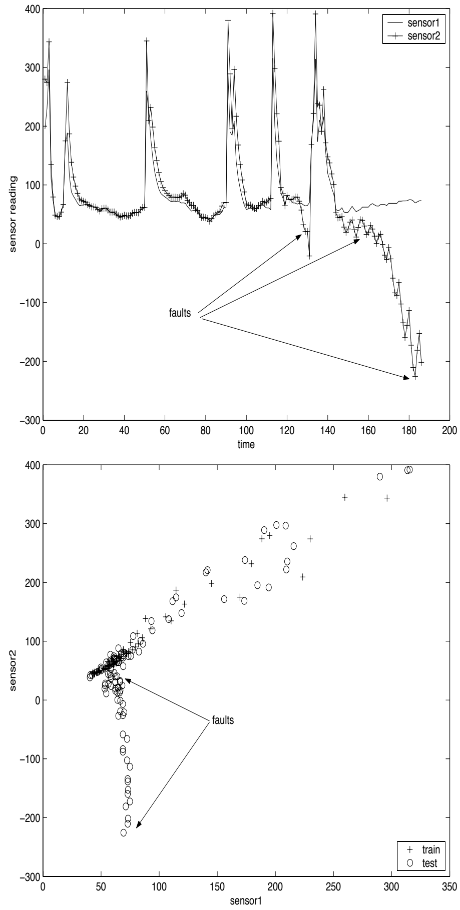

Fig. 2. Sensor readings in time 1:80 are used as training data. There are faults in the test data. Sensor2 drifts off.

The ROC curve can not be produced unless both negative (clean) and positive (faulty) sensor data is available.

In the next section, we give an algorithm to produce the faulty data based on the analyst specified sensor thresholds and the training data.

III. ROC in the Absence of Positive Examples

In order to be able to use the ROC curve, we first set aside a portion of the original training set as the validation set. The training and validation sets (for a specific partition) will be denoted by:

$$X_{train} = \{x'_1, \ldots, x'_{N_{train}}\} \quad (1)$$
$$X_{validation} = \{x''_1, \ldots, x''_{N_{validation}}\} \quad (2)$$
$$X_{test} = \{x_1, \ldots, x_M\} \quad (3)$$

where $N_{train} + N_{validation} = N$ which was the original training set.

The labels T_{train} and $T_{validation}$ are all False, because the training data is clean. The labels for the test set T_{test}

may contain positives. Let the threshold for each input coordinate be $d_i, i = 1, \ldots, k$, where k is the input dimensionality. Let D_i denote a vector with 0 s everywhere except d_i at the ith coordinate.

Given positive thresholds per sensor $d_i \in R, i = 1, \ldots, k$, in order to compare any two classifiers C1 and C2 we use the following algorithm:

Train both classifiers on the training data.

Extend the validation set based on each validation input and sensor thresholds as follows:

for each input coordinate i $(1, \ldots, k)$

for each original validation input x'_j $(1, \ldots, N_{validation})$

$u = x'_j + D_i$

includeOK = true

for each original validation input x'_{jj} $(1, \ldots, N_{validation})$

if $|u - x'_{jj}| < [d_1, \ldots, d_k]$ for all input coordinates

includeOK = false

if includeOK == true, add u and True to the validation set.

$u = x'_j - D_i$

includeOK = true

for each original validation input x'_{jj} $(1, \ldots, N_{validation})$

if $|u - x'_{jj}| < [d_1, \ldots, d_k]$ for all input coordinates

includeOK = false

if includeOK == true, add u and True to the validation set.

Compute the ROC curve for each classifier using the extended validation set.

The classifier with less cost average (with respect to sensors) around the thresholds d_i is better classifier.

The extended validation set contains both negative (from the original validation set) and positive (from the disturbed validation set) inputs. If sensors show variety in terms of likelihood of faults, direction of faults, or if faults come in multiple sensors, the above extension of the validation set should be modified accordingly.

The positives in the extended validation set form an envelope around the validation inputs. (See figure 3 for example.)

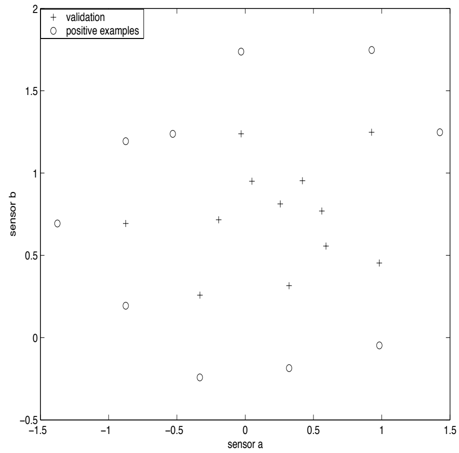

Fig. 3. Positive examples created for hypothetical data.

References

[1] Bradley, A. P. (1997) The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms, *Pattern Recognition*, Vol. 30, No. 7, pp. 1145–1159.

[2] Egan, J. P. (1975)*Signal detection theory and ROC analysis*, New York:Academic Press.

[3] Fawcett, T. (2001) Using Rule Sets to Maximize ROC Performance, *IEEE International Conference on Data Mining (ICDM)*, pp. 131-138.

[4] Hanley, J. A. , McNeil, B. J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, 143, pp. 29-36.

[5] Metz, C. E. (1978) Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, 8, pp. 283-298.

[6] Mozer, M. C., Dodier, R., Colagrosso, M. D., Guerra-Salcedo, C., & Wolniewicz, R. (2002) Prodding the ROC curve: Constrained optimization of classifier performance, *In T. Dietterich, S. Becker, & Z. Ghahramani (Eds.), Advances in neural information processing systems XIV, Cambridge, MA: MIT Press* , pp. 1409-1415.

[7] Provost, F., Fawcett, T. and Kohavi, R. (1998) The case against accuracy estimation for comparing induction algorithms, *Proc. 15th International Conf. on Machine Learning, Morgan Kaufmann, San Francisco, CA*, pp. 445–453.

[8] Provost, F., Fawcett, T. (2001) Robust Classification for Imprecise Environments, *Machine Learning*, vol. 42, no. 3, pp. 203–231.

[9] Zweig, M.H., Campbell, G. (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry*, 39, pp. 561-577.