Anomalies Detection Based on the ROC Analysis using Classifiers in Tactical Cognitive Radio Systems: A survey

Ahmed Moumena
Electronic Department, Saad Dahlab-Blida University, Algeria

Article Info

Article history:

Received Jun 5, 2016
Revised Aug 8, 2016
Accepted August 24, 2016

Keyword:

Anomaly detection theory
Classifier
Evaluation
Jamming signal
ROC Analysis

ABSTRACT

Receiver operating characteristic (ROC) curve is an important technique for organizing classifiers and visualizing their performance in tactical systems in the presence of jamming signal. ROC curves are commonly used to evaluate the performance of classifiers for anomalies detection. This paper gives a survey of ROC analysis based on the anomaly detection using classifiers for using them in research. In recent years have been increasingly adopted in the machine learning and data mining research communities. This survey gives definitions of the anomaly detection theory and how to use one ROC curve, what a ROC curve, when we use ROC curves.

Corresponding Author:

Ahmed Moumena,
Electronic Department,
Saad Dahlab-Blida University, Algeria.
Email: aamoumena@gmail.com

1. INTRODUCTION

This survey presents a novel machine learning approaches to spectrum sensing in collaborative cognitive radio systems in the presence of source signal, jamming signal and noise. Cognitive radio (CR) is a novel technology that allows improving spectrum utilization by enabling opportunistic access to the licensed spectrum band by unlicensed users. This is accomplished through heterogeneous architectures and techniques of dynamic spectrum access (DSA). The CR is defined as a smart wireless communication system that is aware of its environment and is capable to learn from the environment and adapt its transmission parameters, such as frequency, modulation, and transmission power and communication protocols.

An important aspect of a CR is spectrum sensing (SS), which involves a principal task: jamming signal detection. Jamming signal detection refers to the detection of anomalies created by undesirable signals that disrupt or jeopardize the communications between primary users (PUs) and secondary users (SUs), using machine learning algorithms and patterns recognition. This task is important so that the unlicensed users (jammers) do cause interference to licensed users (PUs, SUs) [Lilian, 12].

The received signal by every wideband CR receiver is used by a classifier for detection at a fusion center (FC) to make a global decision about the availability of anomalies/outliers caused by the effects of jammers. In the last few years, ROC graph is appearing more intensely in the domain of machine learning. Actually, their use as a metric to evaluate machine learning algorithms has become necessary. In this survey, the terms ROC analysis, ROC graph and ROC curve are used but the most used is ROC analysis.

One of the earliest adopters of ROC analysis in machine learning [Spackman, 89] was who demonstrated the value of ROC curves in evaluating and comparing algorithms. Recent years have seen an increase in the use of ROC analysis in the machine learning community.

Most books on data mining and machine learning, if they mention ROC curves at all, have only a brief explanation of method. ROC curves are conceptually easy, but there are some non-obvious complexities that arise when they are used in research and development [Fawcett, 03].

2. ANOMALY DETECTION THEORY

Suppose that for some physical measurement a primary user (PU) produces an output signal $r = \{r(t): t \in [0, T]\}$, over a time interval $[0, T]$. Suppose that the signal may have been produced by jamming signal, source signal and ambient noise of type AWGN. There are two possibilities are called the hypothesis H_0 and the hypothesis H_1, respectively, and are commonly written in the compact notation:

First hypothesis: H_0: events: normal observations.

Second hypothesis: H_1: events: anomaly/abnormal observations.

To decide between the first and second hypotheses one might apply a high threshold to the classifier output r and make a decision that the anomalies are present if and only if the threshold is exceeded the threshold value. The engineer is then faced with the practical question of where to fix the threshold so as to ensure that the number of decision errors is small.There are two types of error possible: the error of missing (decide H_0 under H_1 (problem of anomalies is present)) and the error of false alarm (decide H_1 under H_0 (no problem of anomalies is present)). There is always a compromise between choosing a high threshold to make the average number of false alarms small versus choosing a low threshold to make the average number of misses small. To quantify this compromise it becomes necessary to specify the statistical distribution of r under every of the hypotheses H_0 and H_1.

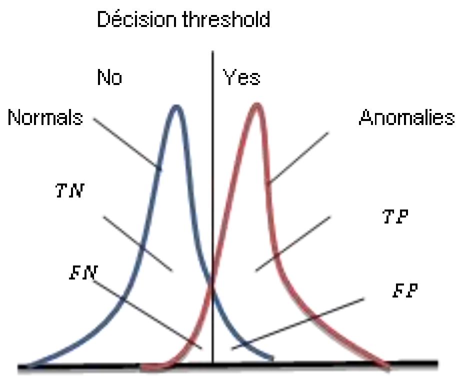

Figure 1. Show Decision Criterion.

3. DEFINITION of ROC CURVE

A ROC curve is a two dimensional (2d) of the accuracy of a classifier for anomaly detection. This 2d curve show, how the true-positive rate (TPR) of detection decreases as the false-positive rate (FPR) increases [Matja, 11]. A ROC plots TPR of detection against FPR. These two types of rate means detection threshold. TPR is highest and FPR is lowest. This principal relation in two components of accuracy will change from one classifier to the next. This makes the form of every ROC different them other. We can use ROC analysis to know how a simple or individual classifier is behaving on dataset?, or to compare the accuracy of two or three or more classifiers on the dataset.

ROC curve explain more detailed analyses about the expected accuracy and cost of the classifier. If we know how abnormal observations events are in relation to normal observations events, we can estimate the ratio of the two kinds of errors for every threshold level, the same case for the costs. We can choice an optimal classifier out of many candidates.

The knowledge of whether a simple ROC dominates all others is important. It is important to imprecise event-class and error-cost. In addition, when there is no simple dominator, the various ROC which share domination can be used to build hybrid classifier which equals or outperforms any simple classifier [Provost, 01.a].

A matrix is used called confusion matrix (represents the confusion between classes). There are four outputs for classification of each instance/observation or pattern. If the instance is positive and is classified as such then we denote it as classifiers in $(0, 0)$ and $(1, 1)$ are called default detectors. The perfect classifier is $(TPR, FPR) = (1, 0)$.

All classifiers are located on the diagonal line have the same performance. It is said they have no information about the problem. All classifiers located above the diagonal are useful. Confusion matrix calculation: that show correct and

Table 1.Confusion Matrix.

Decision	Events Anomaly	Normal
Yes	TP Hit	FP False-alarm
No	FN Miss incorrect predictions	TN Correct-rejection

TP=true positives: an anomaly observation is classified correctly such as anomaly observation, which means present and detected. FP=false positives: a normal observation is classified such as anomaly observation, which means not present but detected. TN=true negatives: a normal observation is classified such as normal observation, which means not presented and not detected. FN=false negatives: an anomaly observation is faults classified such as normal observation, which means present but not detected.

True positive rate: $TPR = \frac{TP}{TP+FN}$: positives correctly classified/total positives

False positive rate (also called false alarm rate): $FPR = \frac{TN}{FP+TN}$: negatives incorrectly classified/total classified.

$P = Precision = \frac{TP}{TP+FP}$: Positives prediction rate that are corrects.

True negative rate: $TNR = \frac{FP}{FP+TN}$

False negative rate: $FNR = \frac{FN}{FN+TP}$.

$$F - mesure = \frac{TP + TN}{TP + FN + FP + TN}$$

Additional terms associated with ROC curves are:

Sensitivity= recall

Specificity= $\frac{true\ negatives}{false\ positives+true\ negatives}$=1- false positive rate.

Positive predictive value=precision.

4. ROC SPACE

ROC graphs are two-dimensional curves in which TPR is plotted on the y-axis and FPR is plotted on the x-axis. An ROC graph depicts relative tradeoffs between benefits (TP) and costs (FP). Figure.2, below shows an ROC graph with five classifiers labeled A through E. A discrete classifier is one that outputs only a class label. Each simple classifier detector produces an (FPR, TPR) pair corresponding to a simple point in ROC space. The classifiers in Figure.2 are all discrete classifiers. Several points in ROC space are important to note. The lower left point $(0, 0)$ represents the strategy of never issuing a positive classification; such a classifier commits no FP errors but also gains no TP. The opposite strategy of unconditionally issuing positive classifications, is represented by the upper right point $(1, 1)$. The point $(0, 1)$ represents perfect classification. D performance is perfect as shown in the Figure.

Informally, one point in ROC space is better than another if it is to the northwest (TPR is higher, FPR is lower). Classifiers appearing on the left-hand side of an ROC graph, near the x-axis, may be thought of as ‘‘conservative’’: they make positive classifications only with strong evidence so they make few false positive errors, but they often have low TPR as well. Classifiers on the upper right-hand side of an ROC graph may be thought of as ‘‘liberal’’: they make positive classifications with weak evidence so they classify nearly all positives correctly, but they often have high FPR. In Figure.2, A is more conservative than B. Many real world domains are dominated by large numbers of negative instances or observations, so performance in the far left-hand side of the ROC graph becomes more interesting.

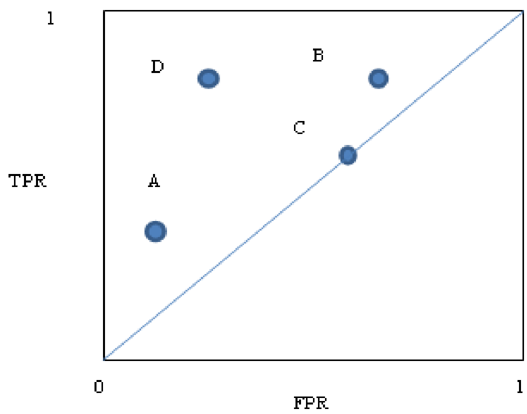

Figure 2. Shows an ROC graph with five discrete classifiers labeled A through E [Fawcett, 03].

5. USING DATASET TEST TO BUILD ROC CURVE OF CLASSIFIERS FOR DETECTION

The classifier output represents all ranges of possible scores.

Points about dataset:

The definition of ‘’event’’ must be clear and must cover all necessary circumstances. Every event ‘’anomaly observations’’ or ‘’normal observations’’ will be ranked independently by classifier detector. Every event should be labeled as ‘’anomaly’’ or as ‘’normal’’. It is necessary to have many examples of both anomalies and normal. Every category of anomalies and normal, there must be representative types and proportions of events.

How to group dataset scores:

Decide what type of threshold to use. We can choose values of thresholds that correspond to fixed levels of FPR. The number of values of thresholds we use will be the number of ROC curve points on the graph. Run the classifier detector on the evaluation dataset. Compare the anomaly classifiers detection result to the ground truth label over all events ‘’anomaly observations: event.1: H_1’’ or ‘’normal observations: even.2: H_0’’.

6. INTERPRETATION OF ROC CURVE

There are two types of events and two types of accuracies possible. ROC curve with two-dimensions (2d), y-axis show success rate (abnormal observations: events) of detection and x-axis show error rate (normal observations: events). Success is better and error is not good. Ideal ROC means in y-axis the values grows at a quickest rate and in x-axis the values rises swiftly upward, the error values for (normal observations: events) x-axis must rise large. The perfect ROC curve touches the point (0, 1). There is different form of ROC curves which means different levels of classifier accuracy. A perfect classifier will have a success rate of 1.0 for (abnormal observations: events) while having an error rate of 0.0 for (normal observations: events). Unfortunately, this result is difficult to obtain.

Each ROC is based on the measurements of classifier performance at different decision threshold values. Based on the ROC curve, the stricter threshold value closer to (0, 0) point and the more lenient threshold value appear closer to (1, 0) point. The (0, 0) point corresponds to tell (NO) and (1, 0) point corresponds to tell (YES). The aim is to minimize expected cost and to maximize the TPR given a fixed FPR.

Finally, we will give the concept of detection which the ROC curve will be easier to understand. Table 1**.** describe the TPR and FPR for four possible detection outputs. There are two possible true classes: (anomaly observation: events) and (normal observation: events) and possible decision classes (YES: means anomaly and NO: it is normal. Two of these outputs are successful when the decision matches truth and two are erroneous, when there is a mismatch between the decision and the truth. We will use the terms TP and FP because they are frequently used in formulas and to represent the axes of ROC curves.

7. ADVANTAGES OF USING ROC ANALYSIS

a. Visualize accuracy of classifier for detection.
b. Facilitate the comparison of more classifiers.
c. Recognize the importance of value threshold decision.

8. MEASURES OF ROC ANALYSIS FOR ANOMALY DETECTION

8.1 Measures of Accuracy

Visualization of ROC curve provides to a classifiers global accuracy. The nature of ROC is steeper which means anomaly observations rate is greater. The nature of ROC is flatter which means the normal observations rate is greater. We can view that ROC curve approaches at the point of perfection (0, 1). Neyman-pearson criterion which means TPR at fixed FPR. The first means that there is a particular FPR, and the second means a simple measure of accuracy. TPR at fixed FPR and AUC will be explained in section.

.

8.1.1 Neyman-Pearson Criterion

The importance of Neyman-Pearson criterion of anomaly detection is to maximize the rate of HIT (TP) at a fixed rate of false-alarms (FP). After FPR is fixed, it remains to know what the best TPR achievable for that level is. It is possible to see at the global figure of the ROC curve, and decide upon a fixed FPR and this in statistical hypothesis may not be correct. The ROC provides essential clue if the ROC is steep in the region of interest. Similarly, if the ROC is very flat in the region of interest, than a larger FPR will not gain much. Using measure of accuracy, comparing two or multiple ROC curves, easy find the ROC curve with greater TPR for a given fixed FPR.

8.1.2 Area under the ROC curve – AUC

To compare classifiers for detection we reduce ROC performance to a simple scalar value representing expected performance. The area of this zone is called the "Area Under Curve or AUC [Bradley, 97], [Hanley, 82] and has become a better alternative of exactitude (accuracy) or error to evaluate the classifiers. Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1.0. However, because random guessing produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, no realistic classifier should have an AUC less than 0.5 [Fawcett, 03]. The AUC of a classifier is equivalent to the probability that a classifier give a higher ranking of a positive element to a negative element. The AUC is also very close to the coefficient of Gini [Breiman, 84] which correspond to the area between the ROC curve and diagonal space. In [David, 11] the relationship between AUC and coefficient Gini was specified to give

$$\mathrm{Gini} + 1 = 2 \times \mathrm{AUC} \qquad (1)$$

Figure 3. with the area under two ROC curves, A and B. In Figure.3a**,** the classifier B has a largest area and therefore best average performance. In Figure 3b shows the AUC of a binary classifier A, and a scoring classifier B. classifier A represents performance of classifier B when B is used with an individual fixed threshold. Through the performance of the two is equal to a given point (A’s threshold), A’s performance becomes inferior to B further from this point. It is possible for a high classifier to perform worse in a specific region of ROC space than a low AUC classifier.

Figure 3a. shows an example of this: classifier B is generally best than A except at FPR>0.6, where A has slight advantage. In practice AUC performs very good and is always used a general measure is desired. Because of its extremely general nature, the AUC measure is ideally suited for high-level classifier comparisons, such as in evaluating core anomaly detector technology. It is also useful for summarizing the entire figure of a classifier’s performance. If you have more specific needs in a particular detection setting, it may be preferable to use the partial-AUC or even an isoperformance line in conjunction with an ROC convex hull to provide more meaningful comparisons.

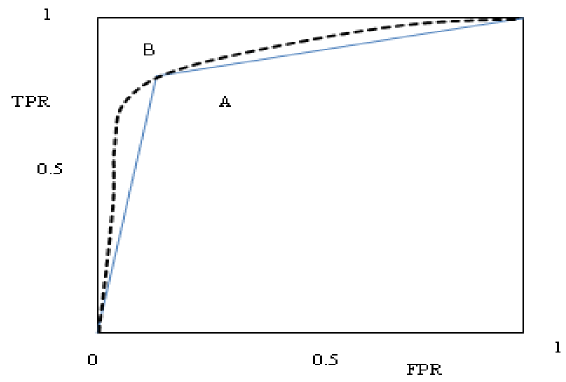

Figure 3a. The curve on the left shows the area under two ROC curves.

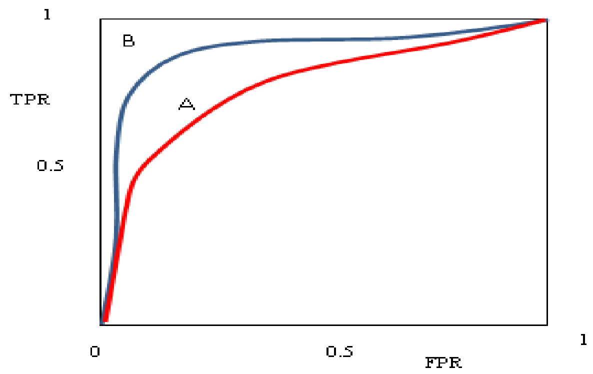

Figure.3b. This curve shows the area under the curves of a discrete classifier (A) and a probabilistic classifier (B).
Figure.3: ROC and AUC curves [Fawcett, 07].

8.1.3 Partial-AUC

Partial area under the curve is just like the global area under the curve, except that only a subset of the ROC picture is considered. For instance, if it is known that a FPR of 0.50 or greater is completely unacceptable, then only the left half of the ROC curve need be considered. In this case, partial area under curve will no longer range between 0.5 and 1.0, like the AUC measure did, but it will have a new minimum and maximum level (from 0.125 to 0.5, respectively) which always depends on how much of the picture is considered.

In order to narrow down the region of the curve of interest, it is necessary to have in mind a fixed maximum FPR, a fixed minimum TRP. In terms of focus it is somewhere between global area and the FPR fixed at TPR, because it considers accuracy over a range of the ROC graphs but not over the total curve picture. However, like the global measure AUC, it suffers from ambiguity because if the curves cross one another within the region of interest, it is not clear that one of the curves having a larger area will unambiguously be the best classifier to use under deployment conditions. However, if a simple ROC curve dominates the region of interest, then the partial AUC measure becomes less problematic [Walter, 05], [Man, 13].

9. EVALUATION OF ANOMALIES DETECTION USING (ROC) OR (AUC)

Standard measures for evaluating anomaly detection problems:

Recall (Detection rate or true positive rate (TPR)) ratio between the number of correctly detected anomalies and the total number of anomalies. False alarm (false positive rate (FPR)) ratio between the number of data records from normal class that are misclassified as anomalies and the total number of data records from normal class. ROC Curve is a tradeoff between detection rate (TPR) and false alarm rate (FPR). Area under the ROC curve (AUC) is calculated using a trapezoid rule.

Main idea: build a classification model for normal (and anomalous) events based on labeled training data, and use it to classify each new unseen event Classification models must be able to handle skewed (imbalanced) class distributions. Use modified classification model to learn the normal behavior and then detect any deviations from normal behavior as anomalous.

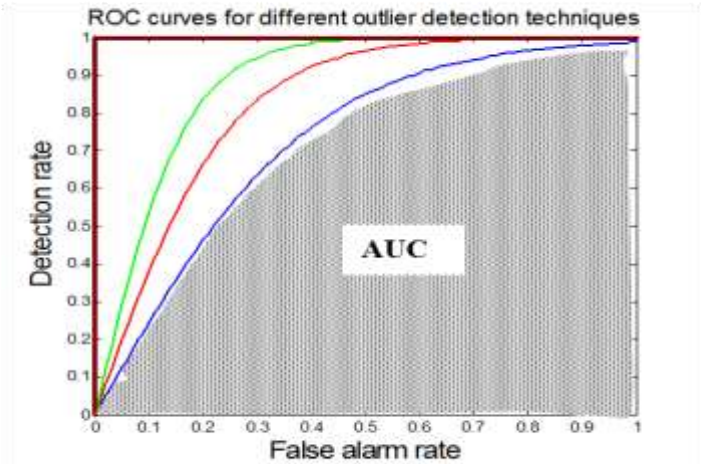

Figure 4. ROC curves for different anomalies detection methods [Arindam, 08].

10. PROPOSITIONS OF ROC ANALYSIS

Dataset must be classifiable in the first category (anomaly observation) and in the second category (normal observation).What is an anomaly/outlier or abnormal observation and what is a normal observation?. All classifiers having threshold values benefit from the bi-dimensional (2d) visualization of the ROC curve. Evaluate the performance of a classification system is a very important issue because these performances can be used for learning as such or to optimize the values of the hyper-parameters of the classifier. For a long time, the criterion used to evaluate this performance was the correct classification rate, that is to say the number of elements in a test database correctly classified. The problem is that such a test is not suitable for ill-defined environments. In many situations, not all errors have the same consequences. Some errors have cost more than others, for example, medical diagnostics. Improper diagnosis or treatment can, in fact, have different costs or dangers according to the type of error. We provide an overview of the evaluation criteria of classification systems in two classes as discussed above and more generally multi-class systems that we will discuss in the sections.

11. GENERALIZATION AND DECISION PROBLEMS OF THE ROC ANALYSIS TO MULTICLASS PROBLEMS

11.1. Multi-class ROC

With more than two classes the situation becomes very complex if the global space is to be managed. The confusion matrix with $n > 2$ classes becomes a matrix with a dimension $(n \times n)$. The n correct classification and $(n^2 - n)$ possible errors. For example for $n = 3$ classes, we get 6 dimensional spaces. In the paper [Srinvasan, 99] has described that the analysis behind the ROCCH extends to multiple classes and multidimensional convex hulls.

In [Provost, 01.b], [Fawcett, 06] and [Landgrebe, 06] proposes to manipulate n classes by generating n ROC curves, one for each class. On the set of all classes, the $i^{th}(i \in \{1, \dots, n\})$ ROC curve corresponds to the evaluation of performances using the class c_i as positive class and all other classes as negative, denoted N_i: $P_i = c_i$

$$N_i = \cup_{j \neq i} c_j \in C \qquad (2)$$

With $i, j \in \{1,2, \dots, n\}$ and C is the set of all classes.

The cost of misclassification is, for this approach, fixed for each class because we do not seek to differentiate the errors. Under these conditions, space performance evaluation is n dimensions, which amounts to use only the elements of the principal diagonal of the confusion matrix. For example, for three classes ($n = 3$), we obtain a three-dimensional space easily representable.

Now, we will position in the context of comparing the performance of classifiers. We need it to compare two hyperplanes. The problem is that according to the areas of the space performance of the classifiers may vary. We can have on one area a hyperplane is better than other and in other area the second hyperplane which is better than the first. This is why in the literature when trying to compare different classification systems; we reduce hyperplanes in to scalar values. In the general case, the scalar value that is used to characterize the performance of ROC multi-class is Volume Under the ROC hyper-surface (VUS).

11.2. Multi-class AUC

The Area Under Curve is a measure of the discriminability of a pair of classes. In a two-class problem, the AUC is a simple scalar value, but a multi-class problem introduces the issue of combining multiple pairwise discriminability values [David, 11].

One approach to calculating multi-class AUCs was taken by [Provost, 01.b] in their work on probability estimation trees. They calculated AUCs for multi-class problems by generating every class reference ROC curve in turn, measuring the AUC, and then summing the AUCs weighted by the reference classs prevalence in the dataset. More precisely, they define

$$AUC_{global} = \sum_{c_i \in C} AUC(c_i)\,(p_i) \qquad (3)$$

Where $AUC(c_i)$ is the area under the class reference ROC curve for c_i, as in equation above. This definition requires only $|C|$ AUC calculations, so its overall complexity is $O(|C|nlogn)$. The advantage of AUC formulation is that $AUC_{global}$is generated directly from class reference ROC curves, and these curves can be generated and visualized easily. The disadvantage is that the class reference ROC is sensitive to class distributions and error costs, so this formulation of $AUC_{global}$is as well. The paper [David, 11] takes a different technique in their derivation of a multi-class generalization of the Area Under Curve. They desired a measure that is insensitive to class distribution and error costs. The derivation is too detailed to summarize here, but it is based upon the fact that the Area Under Curve is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. From this probabilistic form, they derive a formulation that measures the unweighted pairwise discriminability of classes. Their measure, which they call M, is equivalent to:

$$AUC_{global} = \frac{2}{|C|(|C|-1)} \sum_{(c_i,c_j)\in C} AUC(c_i, c_j) \qquad (4)$$

Where n is the number of classes and $AUC(c_i, c_j)$ is the area under the two-class ROC curve involving classes c_i and c_j. The summation is computed over all pairs of distinct classes, irrespective of order. There are $\frac{2}{|C|(|C|-1)}$ such pairs, so the time complexity of their measure is $O(|C|^2nlogn)$. While Hand and Tills formulation is well described and is insensitive to changes in class distribution, there is no easy way to visualize the surface whose area is being calculated.

12. COMPARING MANY CLASSIFIERS FOR ANOMALIES DETECTION

When multiple classifiers are used on the same dataset, we can plot their ROC on the same figure. This facilitates the conclusions about dominance.

12.1 Dominance of ROC curve

In Figure.5, we remark that curve A dominates curves B, C and D completely; means the classifier A outperform the others classifiers B, C and D. curves A, B, C dominate over a select region of the ROC curve.

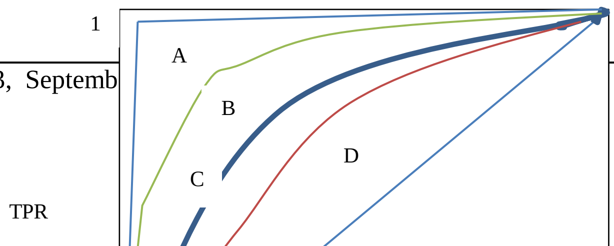

Figure 5. Four ROC curves with different values of the area under curve.

12.2 ROC convex hull (ROCCH)

The ROCCH shows the best possible performance of a set of classifiers, if you take the maximum of accuracy of every classifier and interpolate between different classifiers whenever necessary to correct for any hulls. In the correction, to join two or more classifiers by straight line is an interpolation. The points (0, 0) and (1, 1) can also be used in building the ROCCH. If there are many ROC curves, the best method to compare them is to construct the ROCCH and look which curves dominate over which regions of the figure. We can use the ROCCH to guide the construction of a hybrid classifier which is good when we compare with a simple or individual classifier.

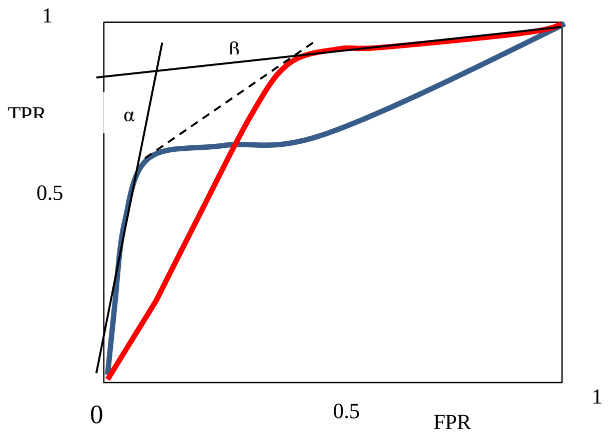

Figure 6. Lines α and β show the optimal classifier under different sets of conditions [Fawcett, 07].

If you aim to cover just 40% of the true positives you should choose method A, which gives a false positive rate of 5%.

If you aim to cover 80% of the true positives you should choose method B, which gives a false positive rate of 60% as compared with A's 80%.

If you aim to cover 60% of the true positives then you should combine A and B.

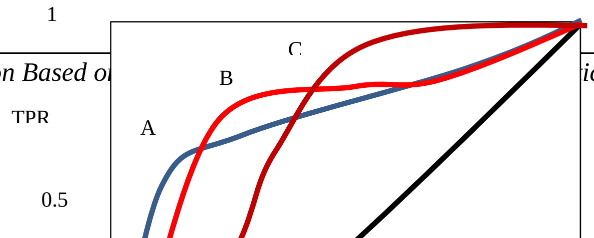

Figure 7. The ROC convex hull identifies potentially optimal classifiers [Fawcett, 07].

12.3 Iso-performance lines

By separating classification performance from class and cost distribution assumptions, the decision goal can be projected onto ROC space for a neat visualization. Formally, let the prior probability of a positive example be $p(\mathrm{p})$, so the prior probability of a negative example is $p(\mathrm{n}) = 1 - p(\mathrm{p})$. Costs of false positive and false negative errors are given by $c(\mathrm{Y}, \mathrm{n})$, and $c(\mathrm{N}, \mathrm{p})$, respectively. The expected cost of a classification by the classifier represented by a point (TP, FP) in ROC space is:

$$p(\mathrm{p}).(1 - TP).c(\mathrm{N}, \mathrm{p}) + p(\mathrm{n}).(1 - FP).c(\mathrm{Y}, \mathrm{n}) \tag{6}$$

Therefore, two points (TP_1, FP_1) and (TP_2, FP_2) have the same performance if:

$$\frac{TP_2 - TP_1}{FP_2 - FP_1} = \frac{p(\mathrm{n})c(\mathrm{Y,n})}{p(\mathrm{p})c(\mathrm{N,p})} \tag{7}$$

This equation defines the slope of an iso-performance line, i.e., all classifiers corresponding to points on the line have the same expected cost. Each set of class and cost distributions defines a family of iso-performance lines. Lines ‘’more northwest’’ having a larger TP intercept are better because they correspond to classifiers with lower expected cost [Provost, 97].

13. COMBINING CLASSIFIERS

Suppose we have generated two classifiers, A and B, which score clients by the probability they will buy the policy.

In ROC space,
$A's$ best point lies at $(0.1, 0.2)$ and
$B's$ best point lies at $(0.25, 0.6)$
We want to market to exactly 800 people so our solution constraint is:
$fp\,rate * 3760 + tp\,rate * 240 = 800$
If we use A, we expect:
$0.1 * 3760 + 0.2 * 240 = 424$ Candidates which is too few.
If we use B we expect:
$0.25 * 3760 + 0.6 * 240 = 1084$ Candidates which is too many.
We want a classifier between A and B.
The solution constraint is shown as a dashed line.
It intersects the line between A and B at C, approximately $(0.18, 0.42)$

A classifier at point C would give the performance we desire and we can achieve it using linear interpolation.

Calculate k as the proportional distance that C lies on the line between A and B:
$k = (0.18 - 0.1) / (0.25 - 0.1) = 0.53$

Therefore, if we sample $B's$ decisions at a rate of 0.53 and $A's$ decisions at a rate of $1 - 0.53 = 0.47$; we should attain $C's$ performance [Fawcett, 07].

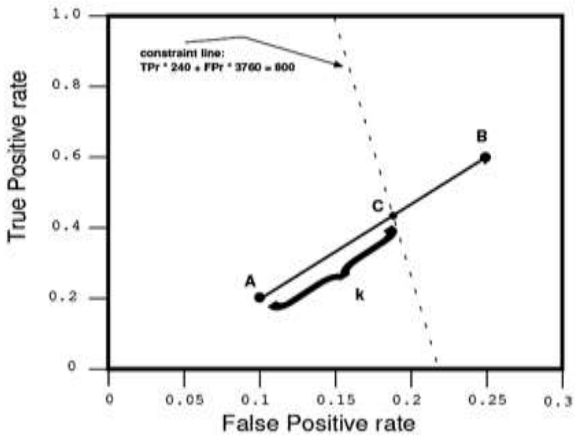

Figure.8: show interpolating classifiers [Fawcett, 07].

14. CONCLUSIONS

This paper has introduced the form and meaning of the ROC analysis and a survey of anomaly/outlier detection using machine learning techniques like one classifier or cooperative of multiple classifiers. Visualize and compare the accuracy of one or more classifiers detectors. Advantages of ROC analysis:

a. Can help to set an ideal value of decision threshold.
b. Can help to evaluate the global rate of errors and global cost of detection.
c. Can help to build an improved accuracy using the collaborative of multiple classifiers for detection.
d. Visualize the accuracy of simple anomaly classifier detection or of many anomaly classifiers for detection.
e. Summarize the global accuracy of individual classifiers.

It can used to show the anomaly detection by using one classifier or multiple classifiers. The ROC analysis can compare many anomaly classifiers using the same dataset test. The ROC analysis can separate the classifiers via the measurement of accuracy. Beyond a simple ROC curve (in general). The data within every category (anomaly observations) is sufficient for the generalization beyond a simple ROC. The dataset text is show the application domain for the categories (anomaly observations). AUC is a better measure than accuracy based on formal definitions of discriminancy and consistency: The paper recommends using AUC as a single number measure to over accuracy when evaluating and comparing classifiers. The ROC convex hull method is a robust efficient solution to the problem of comparing multiple classifiers in imprecise and changing environments.

15. HIGHLIGHTS AND MORE INFORMATION ABOUT ROC ANALYSIS

ROC analysis is vast area of research in artificial intelligence (AI) and remains very interesting domain. In the nineteen fifty, begin to appear and remain to appear in many applications such as: statistics, for classification, estimation and to compute the measures of ROC graph. ROC analysis has continued the progress, and in the literature many books [Thomas, 01], [Vyacheslav, 01], [John, 66], [Egan, 75] have been written about ROC.

Every of these books provide a comprehensive introduction to the theory of signal detection including the use of the ROC curve. [Swets, 00] Provides a good introduction, as well as a case for the increased use of the ROC curve in diagnostic situations. Each successive book also summarizes ideas felt to be important at the time of writing; the later books introduce some topics not present in the earlier books, while the earlier books add important historical context.

ROC graph was defined during the World War Two (WW-II) to help in the detection to identify the enemy ships and planes on the radar. We will develop ROC analysis in Electronic warfare (EW) using anomaly detection theory in the presence of jammers.

REFERENCES

[1] Lilian C Freitas, João Costa, Yomara Pires, Jefferson Morais, Aldebaro Klautau, Data Mining Applied to Cognitive Radio Systems, Computer Science and Engineering, Advances in Data Mining Knowledge Discovery and Applications, intech book, 2012, DOI: 10.5772/51824.
[2] K. A. Spackma, *Signal Detection Theory, Valuable Tools for Evaluating Inductive Learning*, In Proceedings of the Sixth International Workshop on Machine Learning, 1989, pp.160-163.
[3] Fawcett Tom, ROC graphs, Notes and practical considerations for data mining researchers, Tech report, HP Laboratories, Palo Alto, CA, USA, HPL-2003-4.
[4] Egan J. P, Signal Detection Theory and ROC-Analysis, Academic Press, New York, NY, 1975.
[5] Matja Z. Majnik, Zoran Bosni, ROC Analysis of Classifiers in Machine Learning, A Survey. Technical report MM-1, University of Ljubljana, Faculty of Computer and Information Science, 25, Ljubljana, Slovenia, 2011.
[6] Provost F J, Fawcet T, Robust classification for imprecise environments, Machine Learning, 42(3), 2001, pp.203–231.
[7] Bradley A, The use of the area under the roc curve in evaluation of machine learning algorithms, Pattern Recognition, 12, 1997, pp.1145–1159.
[8] Hanley, Mcneil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology, 183, 1982, pp.29-36.
[9] Breiman, and Al, Classification and Regression Trees, Wadsworth International Group, Belmont, CA, 1984.
[10] David J Hand, Robert J Till, A Simple Generalization of the Area Under the ROC Curve for Multiple Class Classification Problems, Machine Learning, 45(2), 2011, pp.171-186.
[11] Fawcett Tom, An introduction to ROC analysis, Pattern Recognition Letters, 2006, 27(8), 861–874.
[12] S. D. Walter: The Partial Area Under the Summary ROC Curve, Statistics in Medicine, John Wiley & Sons, Ltd, 17 May 2005.
[13] Man-Jen Hsu, Huey-Miin Hsueh: The linear combinations of biomarkers which maximize the partial area under the ROC curves, Computational Statistics, 28(2), 2013, pp.647-666.
[14] Arindam Banerjee, Varun Chandola, Vipin Kumar, Jaideep Strivastava, *Anomaly Detection, A Tutorial*, SIAM International Conference on Data Mining (SDM08), 2008.
[15] Srinvasan A: Note on the location of optimal classifiers in n-dimensional ROC space, Technical Report PRG-TR-2-99, Oxford University Computing Laboratory, Oxford, England, 1999.
[16] Provost F., Domingos P: Well-trained PETs: Improving probability estimation trees, CeDER Working Paper #IS-00-04, Stern School of Business, New York University, NY, NY 10012, 2001.
[17] Fawcett Tom: ROC graphs with instance-varying costs, Pattern Recognition Letters, 27(8), 2006, pp.882-891.
[18] Landgrebe T., Duin R, Combining accuracy and prior sensitivity for classifier design under prior uncertainty, Lecture Notes in Computer Science (LNCS), 4109, 2006, pp.512-521.
[19] Provost Foster, Fawcett Tomm, *Analysis and Visualization of Classifier Performance, Comparison under Imprecise Class and Cost Distributions,* In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97), 1997.
[20] Thomas D. Wickens, Elementary Signal Detection Theory, Book, Oxford University press, 276 pages, 2001.
[21] Vyacheslav P. Tuzlukov, Signal Detection Theory, Springer, 725 pages, 2001.
[22] John C. Hancock, Paul A. Wintz, Signal Detection Theory, McGraw-Hill, 247 pages, 1966.
[23] Swets, al Better decisions through science, Scientific American, 283, 200, pp.82-87.