Modifying ROC Curves to Incorporate Predicted Probabilities

C. Ferri\textsuperscript{1}, P. Flach\textsuperscript{2}, J. Hernández-Orallo\textsuperscript{1}, A. Senad\textsuperscript{1}

\textsuperscript{1}Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València
Spain

\textsuperscript{2}Department of Computer Science
University of Bristol
U.K.

\textit{ROCML’05, Bonn, August 11, 2005.}
Outline

- Motivation
- ROC Analysis
- PAUC measure
- pROC Curves
- Properties and Applications of pROC Curves
- Conclusions and Future Work
Motivation

• Cost-sensitive Learning is a more realistic generalisation of predictive learning:
  – Costs are not the same for all kinds of misclassifications.
  – Class distributions are usually unbalanced.

• Receiver Operating Characteristic (ROC) Analysis:
  – Useful for choosing classifiers when costs are not known in advance.

• AUC (Area Under the ROC Curve):
  – A simple measure for each classifier, which estimates:
    • The quality of the classifier for a range of class distributions.
    • A measure of how well the classifier ranks examples (equivalent to the Wilcoxon statistic)
• Applications:
  – ROC analysis and AUC-related measures have been used in many areas: medical decision making, marketing campaign design, probability estimation, etc.

• Problems:
  – AUC ignores the probability values, and it only takes the order into account
  – Other measures (MSE, LogLoss) consider how well the probabilities are calibrated, but not its order.

• Goal:

  Modify ROC analysis in order to consider the probabilities of the instances as well as the order
ROC Analysis

• ROC Analysis is useful when we don’t know in learning time:
  – The proportion of examples of each class (class distribution).
  – The cost matrix.

• ROC Analysis can be applied in these situations. Provides tools to:
  – Distinguish classifiers that can be discarded under any circumstance (class distribution or cost matrix).
  – Select the optimal classifier once the cost matrix is known.
ROC Analysis

• Given a confusion matrix:

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>No</td>
<td>10</td>
<td>40</td>
</tr>
</tbody>
</table>

Predicted

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0.75</td>
<td>0.33</td>
</tr>
<tr>
<td>No</td>
<td>0.25</td>
<td>0.67</td>
</tr>
</tbody>
</table>

• We can normalise each column

ROC diagram
Given several classifiers:

- We can construct the convex hull of their points (FPR, TPR) and the trivial classifiers (0,0), (1,1), (1,0).
- The classifiers falling under the ROC curve can be discarded.
- The best classifier of the remaining classifiers can be chosen in application time.
• If we want to select *one* classifier:

We calculate the Area Under the ROC Curve (AUC) of all the classifiers and choose the one with greatest AUC.
• Wu and Flach method to compute AUC
  – for each positive on the sorted list, accumulate how many negatives follow it.
  – Alternatively, we can accumulate for each negative how many positives precede it

\[
\text{AUC} = 0.8333
\]
AUC

- AUC completely ignores the probability values, it only takes order into account

Small changes in probabilities can give rise to large changes in AUC
pAUC

- We can incorporate probabilities into AUC:
  
  \[
  p\text{GINI} = \sum_{x \in \oplus} \frac{P(x)}{\text{Pos}} - \sum_{y \in \Theta} \frac{P(y)}{\text{Neg}}
  \]
  
  \[
  p\text{AUC} = \frac{\sum_{x \in \oplus} P(x) - \sum_{y \in \Theta} P(y) + 1}{2} = \frac{\sum_{x \in \oplus} P(x)}{\text{Pos}} + \frac{\sum_{y \in \Theta} 1 - P(y)}{\text{Neg}}
  \]

- \(P(x)\) and \(P(y)\) are the predicted probability of positive instance \(x\) and negative instance \(y\) to belong to the positive class
- \(\text{Pos}\) and \(\text{Neg}\) are the total numbers of positive and negative examples
pROC curves

- Has pAUC has a representation in the form of a curve?
  - We study a method based on a representation of segments (with a uniform or normal distribution).
  - The key idea is to consider probabilities as intervals of a fixed width \( d \).
pROC curves

• which is the width \( d \) to use?

  • If \( d = 0 \) \( \Rightarrow \) Area\((d) = \) AUC

  • If \( d = 1.596 \) \( \Rightarrow \) Area\((d) = \) pAUC

  • If \( d \to \infty \) \( \Rightarrow \) Area\((d) \to 0.5 \)
Properties of pROC curves

• Which are the theoretical properties of pROC curves?
  – For any set of examples labelled by probabilities and true labels, if \( d = 0 \Rightarrow \text{Area}(d) = \text{AUC} \)

  – For any set of examples labelled by probabilities and true labels, if \( d \to \infty \Rightarrow \text{Area}(d) \to 0.5 \)

  – \( \text{Area}(d) \) is continuous

  – For any set of examples labelled by probabilities and true labels, there exists a real number \( d \geq 0 \) such that \( \text{Area}(d) = \text{pAUC} \)
Properties of pROC curves

- We have shown that there exists a $d$ such that $\text{Area}(d) = \text{pAUC}$.
- However, is this $d$ unique?
- Consider the ranking $\{1+, 1+, 0.6-, 0.6-, 0.50001+, 0.49999-, 0.45+, 0.45+, 0-, 0-\}$. The AUC of this ranking is 0.68. The pAUC is approximately 0.67.
A more natural option is to consider a normal (Gaussian) distribution centered on the probability. The factor to gauge is also called $d$, but represents the double of the standard deviation. The results are similar (both theoretically and practically). Additionally some of the curves are expected to be “smoother.”
Interpretation of pROC curves

- They smooth the ROC curves, especially when differences between probabilities of consecutive examples of different class are small.
- The value is found globally, so it can give high overlap in some areas but small overlap in others.
- The greater the number of examples, the pROC curve has a closer (but smoother) correspondence to the original ROC curve.
- The differences are important when ROC curves are not very reliable: few examples and small differences in probabilities.
We can employ pROC curves for establishing thresholds according to some application skew. Given the skew at application time, we get the slope of an iso-accuracy line, which gives us the best threshold (or the threshold interval).
pROC curves for establishing thresholds

- if we use the pROC curves, we have a different picture:

<table>
<thead>
<tr>
<th>Skew (Slope)</th>
<th>FPR/TPR ROC curve</th>
<th>Threshold in ROC curve</th>
<th>FPR/TPR in pROC curve</th>
<th>Threshold in pROC using the original prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>60°</td>
<td>0% / 50%</td>
<td>[0.9, 0.8]</td>
<td>9% / 19%</td>
<td>[0.9, 0.8]</td>
</tr>
<tr>
<td>45°</td>
<td>33% / 100%</td>
<td>[0.6, 0.3]</td>
<td>40% / 58%</td>
<td>[0.8, 0.6]</td>
</tr>
<tr>
<td>30°</td>
<td>33% / 100%</td>
<td>[0.6, 0.3]</td>
<td>76% / 95%</td>
<td>[0.3, 0.2]</td>
</tr>
</tbody>
</table>

- The FPR/TPR are much more gradual in the pROC curve than for the original ROC curve.
Conclusions

- AUC has been promoted as a more suitable alternative for evaluating classifiers

- AUC is only concentrated in how the model ranks the examples

- pAUC contemplates the two aspects (ranking and probabilities)

- pROC curves as a graphical representation of classifier performance.
Conclusions and Current Work

- pROC curves help to know different details about the performance of the evaluated model wrt classical ROC curves do.

- Application of pROC curves for establishing classification thresholds. In this context, we are performing experiments comparing pROC and ROC curves.

- Compare theoretically and experimentally (as a selection measure) pAUC wrt other measures sAUC, MSE, LogLoss..