Canada Border Services Agency | Agence des services frontaliers du Canada

Applying Cost Curves to Marine Cargo Container Inspection

DIMACS / DyDAn / LPS Workshop on Port Security, Safety, Inspection, Risk Analysis and Modeling

Rutgers University
November 18th, 2008

Richard Hoshino

Laboratory and Scientific Services Directorate
Canada Border Services Agency

Canada

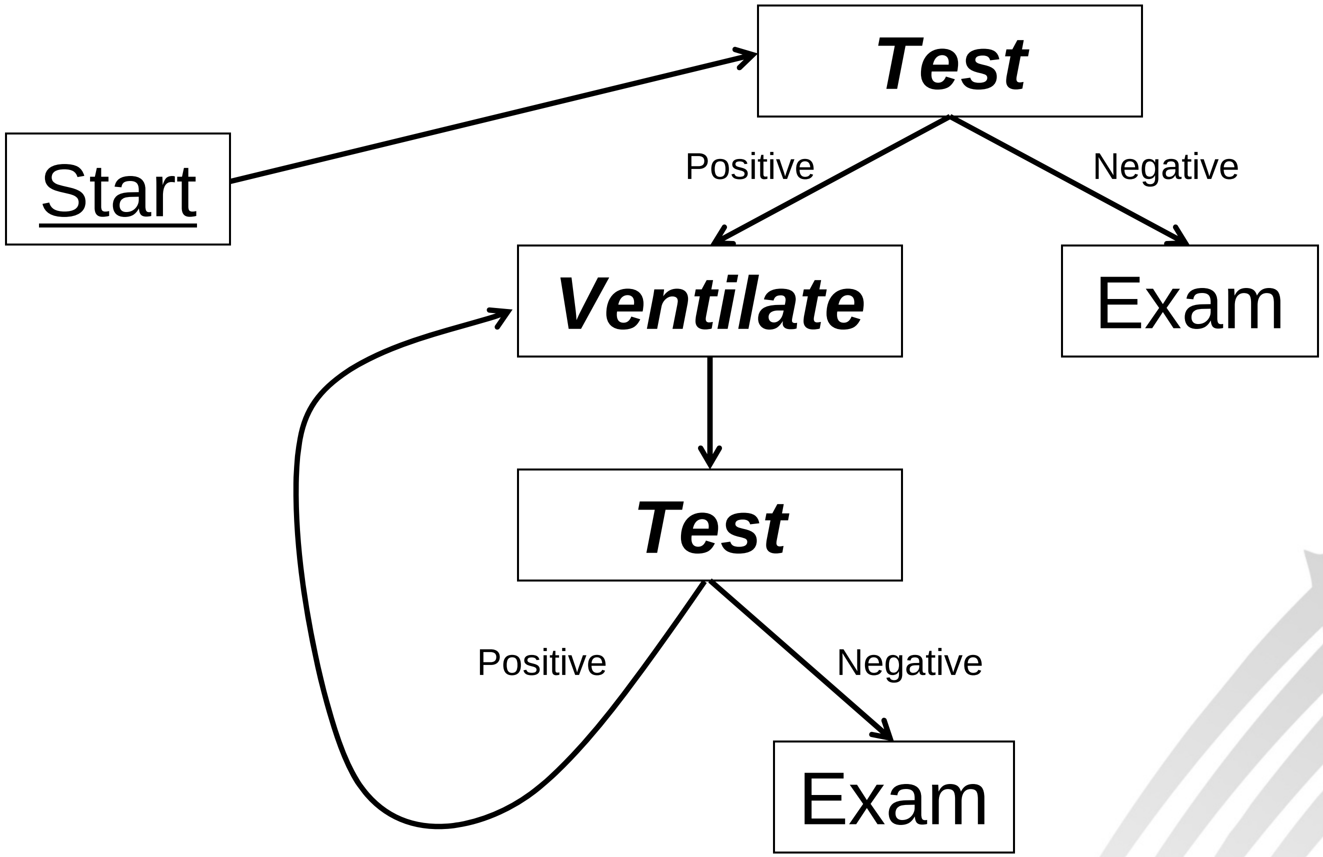
Context

- Each year in Canada, approximately 20,000 marine containers are *referred* for a full examination.

- Some of these containers have been fumigated with chemical compounds to kill invasive alien species.

- If these marine containers are not ventilated properly, fumigants may pose a risk to the health and safety of border service officers.

Flowchart of Current Process



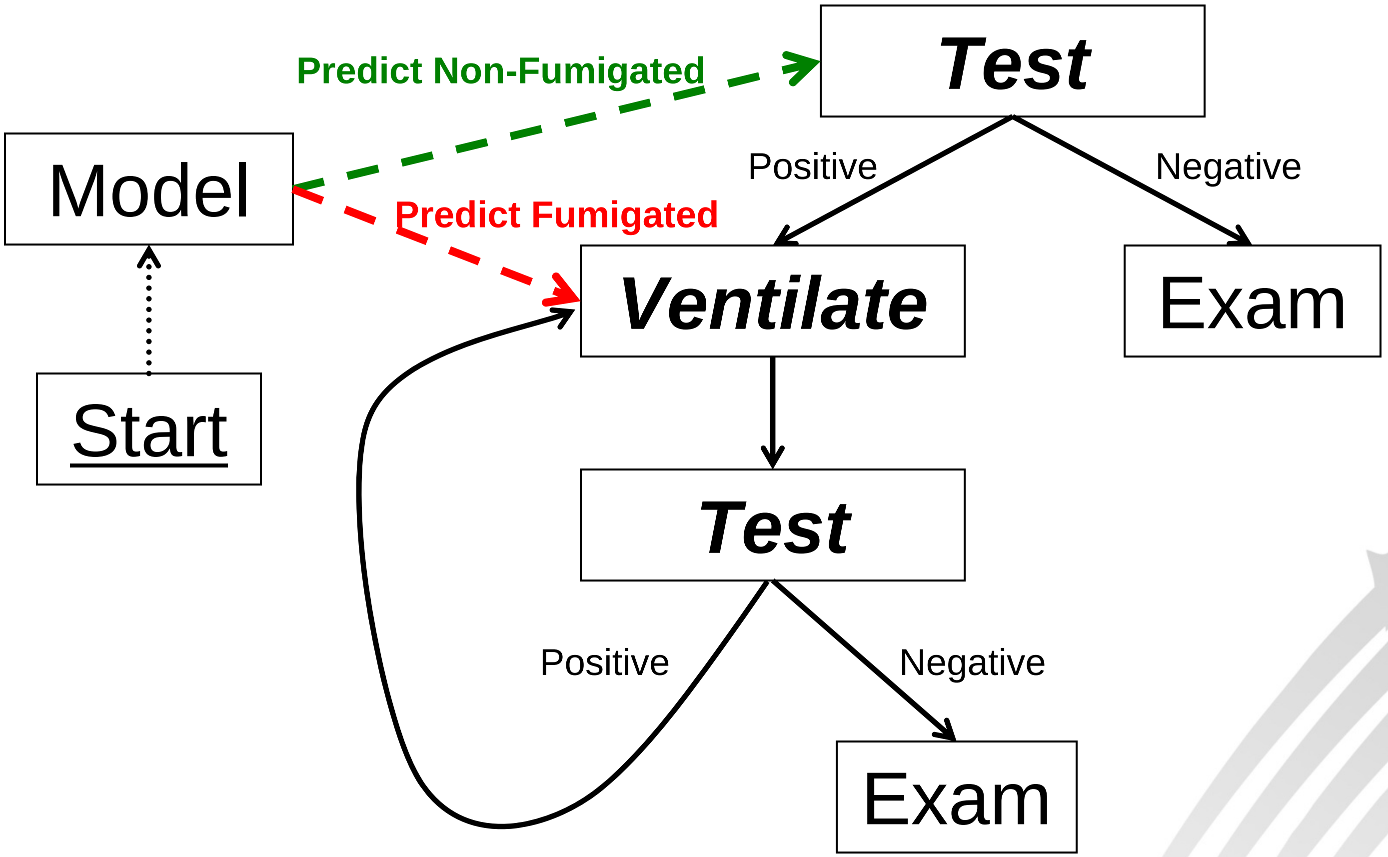
A Simple Yet Powerful Insight

- We can create a mathematical model that predicts whether a container has been fumigated.

- For containers predicted to have been fumigated, we ventilate *prior to* testing.

- Deploying a reliable binary classifier would reduce the overall costs of inspection, creating a more efficient and effective port.

Flowchart of Proposed Process



Predict Non-Fumigated: **Test** first
Predict Fumigated: **Ventilate** first

Misclassification Cost

- The misclassification cost of the Status Quo is

  $M_1 = \mathbf{\#P \times \$C^-}$

- The misclassification cost of the Binary Classifier is

  $$\begin{aligned} M_2 &= \#FN \times \$C^- + \#FP \times \$C^+ \\ &= (\mathit{FNR} \times \#P) \times \$C^- + (\mathit{FPR} \times \#N) \times \$C^+ \\ &= \mathbf{(1 - \mathit{TPR}) \times \#P \times \$C^- + \mathit{FPR} \times \#N \times \$C^+} \end{aligned}$$

- Given a predictive model, its *optimal* binary classifier is the classifier that minimizes the misclassification cost M_2.

The Improvement Curve

- We introduce the **improvement curve**, inspired by the theory of cost curves (Drummond & Holte, 2000).

- Improvement curves measure a model's performance
  - Over <u>all</u> possible class distributions (**#P** vs. **#N**)
  - Over <u>all</u> possible misclassification costs (**\$C⁻** and **\$C⁺**)

- Define the *improvement* to be $\mathbf{I = (M_1 - M_2) \div M_1}$.

Definition of *x*-axis and *y*-axis

- The *x*-axis of the improvement curve is the following expression, denoted as *probability times cost*:

$$x = PC(+) = (\#P \times \$C^{-}) \div (\#P \times \$C^{-} + \#N \times \$C^{+}).$$

- The *y*-axis is the *improvement*, the percentage reduction in misclassification cost by replacing the status quo with the model's optimal classifier:

$$\begin{aligned} y = I(x) &= (M_1 - M_2) \div M_1 \\ &= TPR - [\, FPR \times (\#N \times \$C^{+}) \div (\#P \times \$C^{-}) \,]. \end{aligned}$$

It is straightforward to show that $0 \le x \le 1$ and $0 \le y \le 1$.

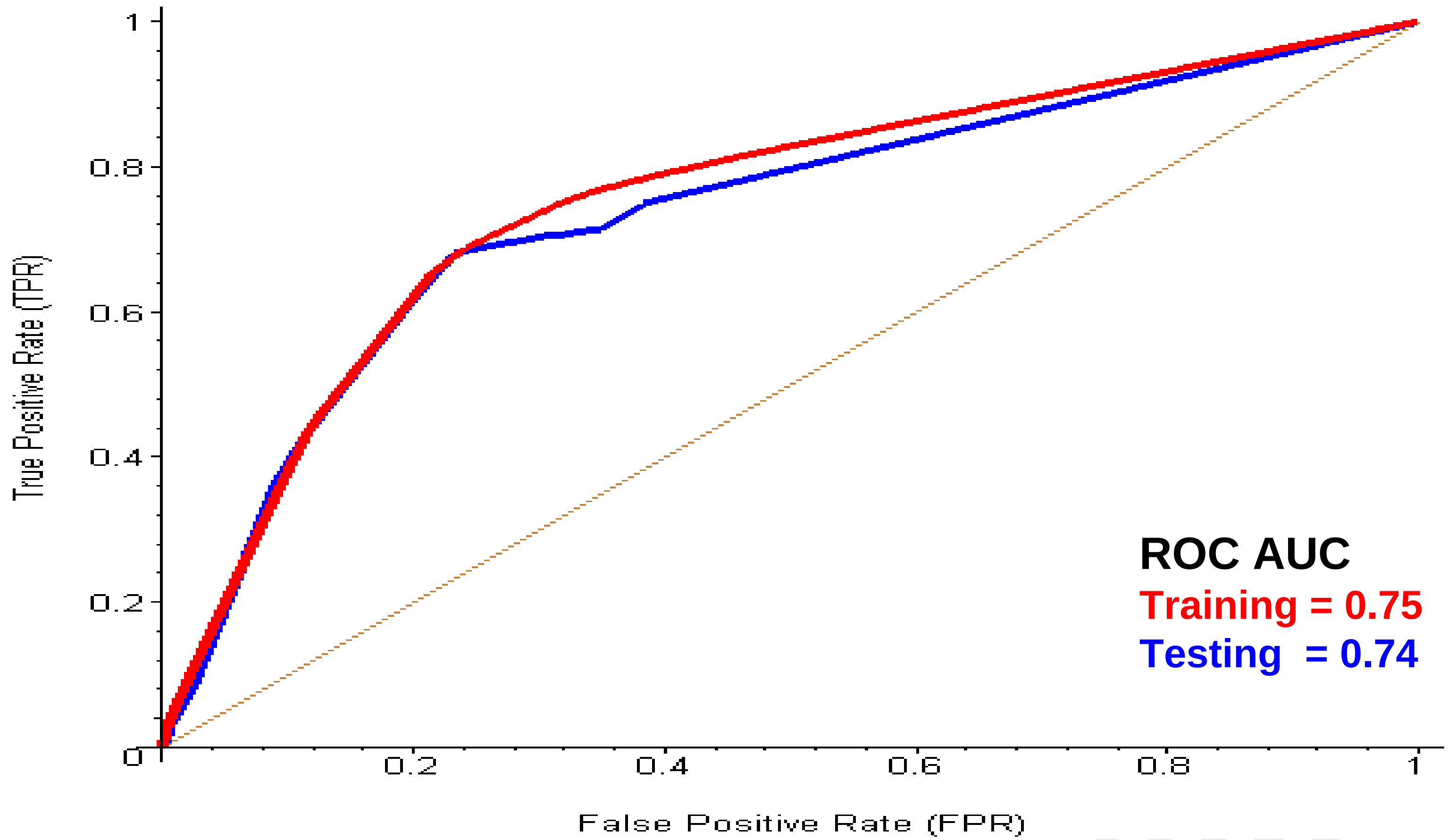
Illustrating the Theory

- We built a simple predictive model based on a 4,200 container data set. The model's four features are:
  - Origin Country
  - Canadian Port of Arrival
  - HS section (e.g. Section 5 = Mineral Products)
  - HS chapter (e.g. Chapter 26 = Ores, Slag, Ash)

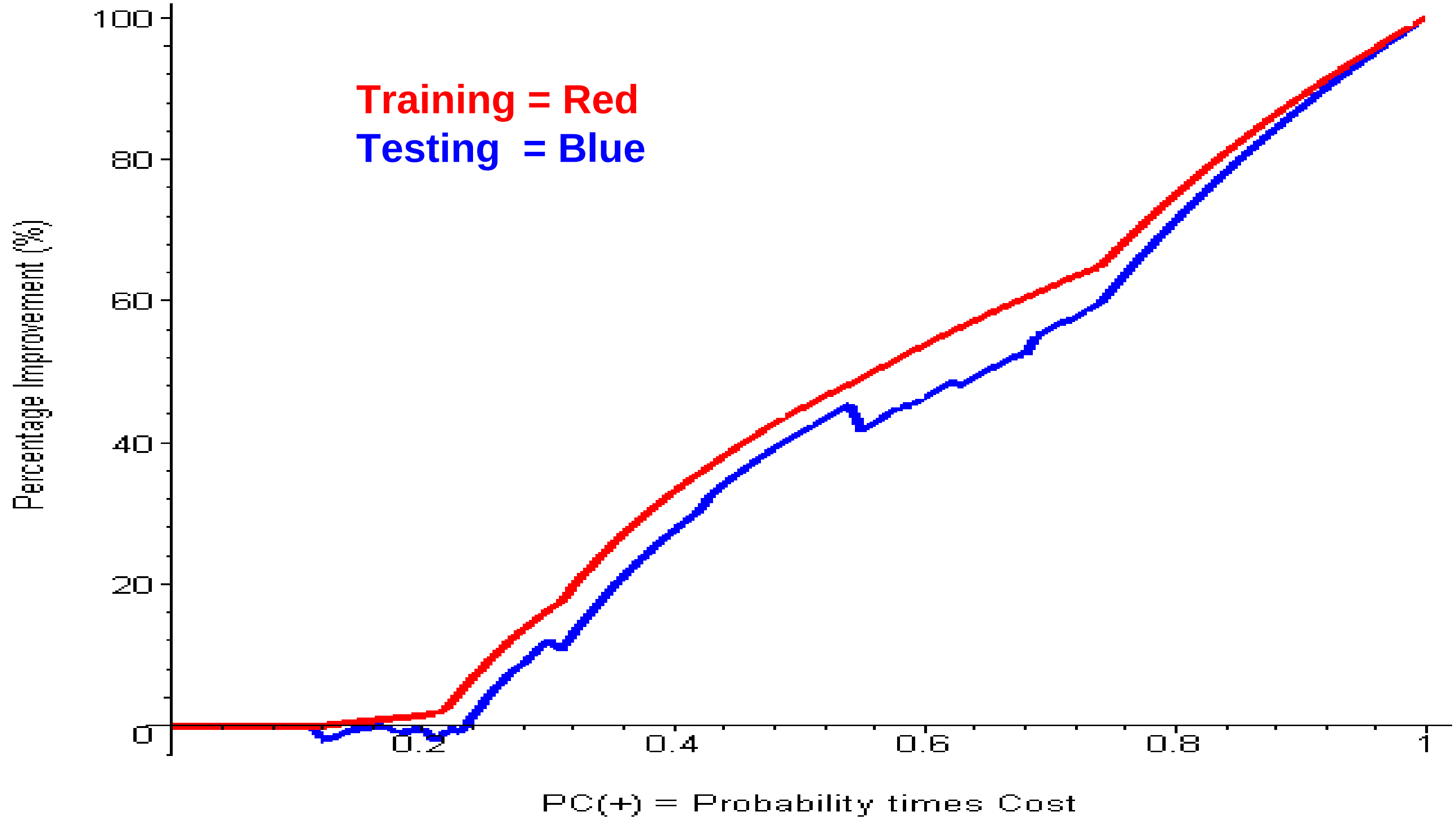
- The model consists of 2^4=16 disjoint classes.

- The data was split 70/30 for Training/Testing.

ROC Curve of Model



Improvement Curve of Model

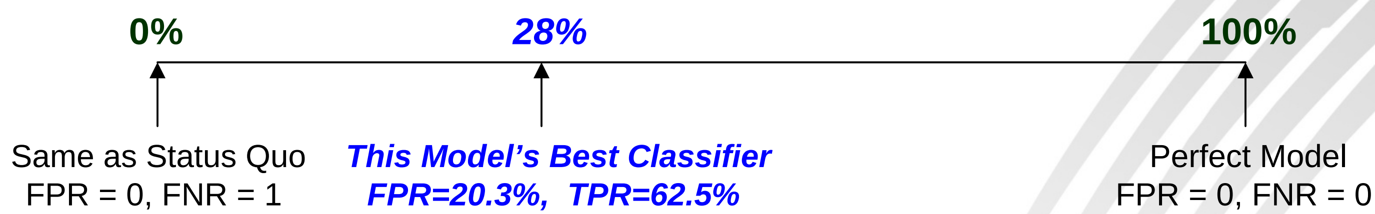


Improvement Curve Interpretation

- Suppose **#N ÷ #P** = 6 and **$C^{-} \div$ C^{+} = 4**. Then,
  x = PC(+) = (#P × C^{-}) ÷ (#P × C^{-} + #N × C^{+}) = 0.4.

- From the improvement curve, we have **y = 28%** (Reading from the Testing Set).

- This simple 4-feature model would have reduced our misclassification cost by 28%.



Conclusion

- The Improvement Curve does the following:
  - Addresses the limitations of ROC curves and the ROC AUC.
  - Measures performance over all possible values of PC(+).
  - Determines a simple condition for when the status quo should be retained.
  - Compares the optimal classifiers of two predictive models.

- The Improvement Curve is an evaluation metric that
  - Is extremely accessible to a non-specialist.
  - Has numerous applications to operations research beyond marine container inspection.