Uplift Model Evaluation with Ordinal Dominance Graphs

Brecht Verbeken, Marie-Anne Guerry & Sam Verboven

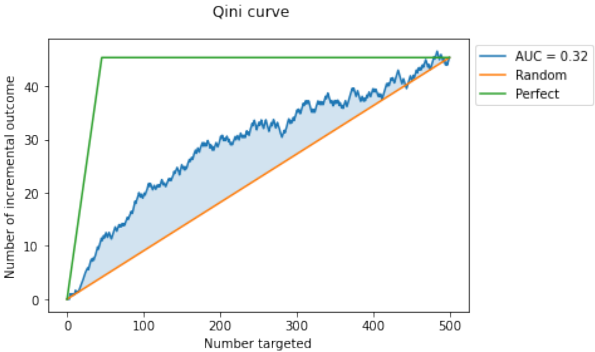
September 19, 2022

VRIJE
UNIVERSITEIT
BRUSSEL

Agenda

1. Qini: How to incorporate information
2. ROCini: How to evaluate
   1. Ordinal Dominance Graphs
   2. Bonus: confidence intervals
3. To evaluate the evaluation metrics

Qini [1]: a recap



$$Q\left(\varphi\right) = \frac{n_T^1(\varphi)}{n_T} - \frac{n_C^1(\varphi)}{n_C}$$

Some remarks about Qini

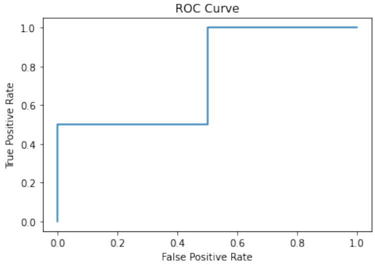
Advantages

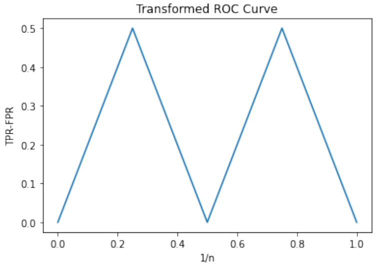
- Visual information
- Intuition: multiplicative and additive counting

Disadvantages

- General inconveniences
- Endpoint is not fixed
- Both terms can compensate for each other ("averaging")
- Confidence intervals: not the most suited object

Back to the basics: ROC and TPR-FPR





Identification: ROC and Qini

$$Q\left(\varphi\right) = \frac{n_T^1(\varphi)}{n_T} - \frac{n_C^1(\varphi)}{n_C}$$

Identification: ROC and Qini

$$Q\left(\varphi\right) = \frac{n_T^1(\varphi)}{n_T} - \frac{n_C^1(\varphi)}{n_C}$$

$$\frac{n_T^1(\varphi)}{n_T} \leftrightarrow TPR(\varphi)$$

$$\frac{n_C^1(\varphi)}{n_C} \leftrightarrow FPR(\varphi)$$

ROCini: a different way of measuring

$$Q\left(\varphi\right) = \frac{n_T^1(\varphi)}{n_T} - \frac{n_C^1(\varphi)}{n_C}$$

$$ROCini\left(\varphi\right) = \left(\frac{n_T^1(\varphi)}{n_T^1} - \frac{n_T^0(\varphi)}{n_T^0}\right) + \left(\frac{n_C^0(\varphi)}{n_C^0} - \frac{n_C^1(\varphi)}{n_C^1}\right)$$

ROCini: a different way of measuring

$$Q(\varphi) = \frac{n_T^1(\varphi)}{n_T} - \frac{n_C^1(\varphi)}{n_C}$$

$$ROCini(\varphi) = \left(\frac{n_T^1(\varphi)}{n_T^1} - \frac{n_T^0(\varphi)}{n_T^0}\right) + \left(\frac{n_C^0(\varphi)}{n_C^0} - \frac{n_C^1(\varphi)}{n_C^1}\right)$$

- T: high $\frac{n_T^1(\varphi)}{n_T^1}$, low $\frac{n_T^0(\varphi)}{n_T^0}$
- C: high $\frac{n_C^0(\varphi)}{n_C^0}$, low $\frac{n_C^1(\varphi)}{n_C^1}$

ROCini: a different way of measuring

$$Q\left(\varphi\right) = \frac{n_T^1(\varphi)}{n_T} - \frac{n_C^1(\varphi)}{n_C}$$

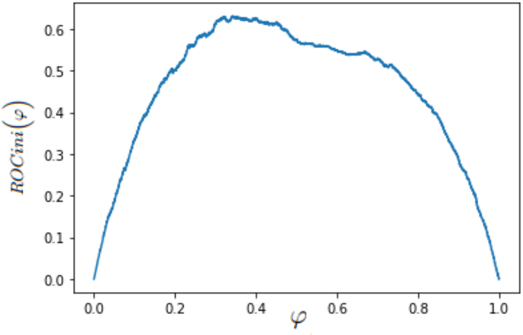
$$ROCini\left(\varphi\right) = \left(\frac{n_T^1(\varphi)}{n_T^1} - \frac{n_T^0(\varphi)}{n_T^0}\right) + \left(\frac{n_C^0(\varphi)}{n_C^0} - \frac{n_C^1(\varphi)}{n_C^1}\right)$$

- T: high $\frac{n_T^1(\varphi)}{n_T^1}$, low $\frac{n_T^0(\varphi)}{n_T^0}$
- C: high $\frac{n_C^0(\varphi)}{n_C^0}$, low $\frac{n_C^1(\varphi)}{n_C^1}$

Ranking: T according P_T^1, C according to $1 - P_C^1 \implies$ "average ranking" according to $1 - P_C^1 + P_T^1 = 1 + ITE$

ROCini

Figure: Example of a ROCini curve



Bridge the gap: Ordinal Dominance Graphs

Definition ([2])

For an arbitrary $t \in]-\infty, \infty[$ we can define $G(t)$ as follows:

$$G(t) = \big(X(t), Y(t)\big) = \big(\mathbb{P}(X \leq t), \mathbb{P}(Y \leq t)\big)$$

$ODG(X, Y)$: all points $G(t)$

Bridge the gap: Ordinal Dominance Graphs

Definition ([2])

For an arbitrary $t \in]-\infty, \infty[$ we can define $G(t)$ as follows:

$$G(t) = \big(X(t), Y(t)\big) = \big(\mathbb{P}(X \leq t), \mathbb{P}(Y \leq t)\big)$$

$ODG(X, Y)$: all points $G(t)$

Theorem ([2])

The ROC-curve is an ODG with $\mathbb{P}(X \leq t) = FPR(t)$ and $\mathbb{P}(Y \leq t) = TPR(t)$

pROCini: best of both worlds

$$pROCini(X(\varphi), Y(\varphi)) = \left(\frac{\frac{n_T^1(\varphi)}{n_T^1} + \frac{n_C^0(\varphi)}{n_C^0}}{2}, \frac{\frac{n_T^0(\varphi)}{n_T^0} + \frac{n_C^1(\varphi)}{n_C^1}}{2} \right)$$

pROCini: best of both worlds

$$pROCini(X(\varphi), Y(\varphi)) = \left(\frac{\frac{n_T^1(\varphi)}{n_T^1} + \frac{n_C^0(\varphi)}{n_C^0}}{2}, \frac{\frac{n_T^0(\varphi)}{n_T^0} + \frac{n_C^1(\varphi)}{n_C^1}}{2} \right)$$

Qini-style

$$A(X, Y) = \int (Y(t) - X(t))dt$$

ODG-style

$$A(X, Y) = \int Y(t)X'(t)dt$$

Confidence Intervals

Confidence bounds for AUROC values are suited for ODG! [3, 4, 5]

$$\left[A(X,Y) - z_{\frac{\alpha}{2}} \times Var(A(X,Y))\,,\, A(X,Y) + z_{\frac{\alpha}{2}} \times Var(A(X,Y))\right]$$

Binormal model [2]

$$X \sim N(\mu_X, \sigma_X^2) \quad \& \quad Y \sim N(\mu_Y, \sigma_Y^2)$$
$$Y - X \sim N(\mu_Y - \mu_X, \sigma_X^2 + \sigma_Y^2)$$

Binormal model [2]

$$X \sim N(\mu_X, \sigma_X^2) \quad \& \quad Y \sim N(\mu_Y, \sigma_Y^2)$$

$$Y - X \sim N(\mu_Y - \mu_X, \sigma_X^2 + \sigma_Y^2)$$

$$A(X, Y) = \phi\left(\frac{\mu_Y - \mu_X}{\sqrt{\sigma_Y^2 + \sigma_X^2}}\right)$$

Simulation Study: compare metrics

Algorithm 1 Simulation

```
Input:
   r: the number of runs
   N: the total number of individuals in the population
   α: the expectation of the probabilities for the control group
   v: the variance of the individual uplifts
   ϵ: the variance of the Gaussian error

Initialize:
   probabilities of positive outcome in control group (PC) ∼ N(α, 0.1);
   individual uplifts (IU) ∼ N(0, M);
   individual uplifts with Gaussian error IU + (IUn) ∼ N(0, ϵ)
   Clamp PC: PC ∈ [0, 1]
   Clamp IU: PC+IU ∈ [0, 1]
   Clamp IUn: PC+IUn ∈ [0, 1]

for i = 1 to r do
   for k = 1 to N do B(0.5) (Obs)            ▷ Bernoulli experiment
      if Obs = 0 then do experiment B(PC) (Out)
      else
         do experiment B(PC+IU) (Out)
      end if
   end for
   for metric in [Qini,ROCini,pROCini] do
      metric(Obs, Out, IU)                   ▷ score perfect model
      metric(Obs, Out, IUn)                  ▷ score model with error
   end for
end for
```

Results

ε	**Qini**	**ROCini**	**pROCini**	**p-value**	**McNemar**
0.05	56.00	56.19	56.19	0.26(0.007)	0.25(0.009)
0.1	67.76	69.97	69.97	< 0.0001	< 0.0001
0.15	77.7	82.44	82.45	< 0.0001	< 0.0001
0.20	83.14	88.82	88.82	< 0.0001	< 0.0001
0.25	87.66	93.43	93.42	< 0.0001	< 0.0001
0.30	91.81	96.87	96.88	< 0.0001	< 0.0001

Table: Performance table of the Qini, ROCini and pROCini based on the McNemar test for $r = 100000$, $N = 500$, $\alpha = 0.5$ and $v = 0.2$.

References I

Nicholas Radcliffe.
Using control groups to target on predicted lift: Building and assessing uplift model.
Direct Marketing Analytics Journal, pages 14–21, 2007.

Donald Bamber.
The area above the ordinal dominance graph and the area below the receiver operating characteristic graph.
Journal of mathematical psychology, 12(4):387–415, 1975.

Simon J Mason and Nicholas E Graham.
Areas beneath the relative operating characteristics (roc) and relative operating levels (rol) curves: Statistical significance and interpretation.

References II

Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 128(584):2145–2166, 2002.

James A Hanley, Barbara J McNeil, et al.
A method of comparing the areas under receiver operating characteristic curves derived from the same cases.
Radiology, 148(3):839–843, 1983.

Corinna Cortes and Mehryar Mohri.
Confidence intervals for the area under the roc curve.
Advances in neural information processing systems, 17, 2004.