Cost-sensitive Learning for Utility Optimization in Online Advertising Auctions

Flavian Vasile (Criteo)
Damien Lefortier (Facebook)
Olivier Chapelle (Google)

Agenda

- **Context**
- Online & Offline Metrics
- Utility Optimization
- Online & Offline Results

Context (1)

- Online Advertising Auctions for Display Advertising; 4 types of players:
  - The auction house: RTB platform,
  - Demand: the advertiser,
  - Supply: the publisher,
  - Bidder: Criteo.

Context (2)

- Most people optimize for deep-funnel events and use a **conversion rate (CR)** prediction model. We focus on this aspect here.

$$
\begin{array}{rl}
y \times v - c & \text{if } p \times v > c \\
0 & \text{otherwise}
\end{array}
$$

Agenda

- Context
- **Online & Offline Metrics**
- Utility Optimization
- Online & Offline Results

Online Metrics

- Conversions are different (e.g., sock vs. car) so we need to weight them by (some flavor of) CPA = Cost / #Conversions.

Offline Metrics

$$\begin{cases} y \times v - c & \text{if } p \times v > c \\ 0 & \text{otherwise} \end{cases}$$

$$\textbf{MSEW} = \frac{1}{N} \sum_i ((y_i - p(\mathbf{x}_i)) \cdot v_i)^2$$

$$\textbf{Utility} = \sum_i \int_0^{p(\mathbf{x}_i)v_i} (y_i \cdot v_i - \tilde{c}) \Pr(\tilde{c} \mid c_i) d\tilde{c}$$

Agenda

- Context
- Online & Offline Metrics
- **Utility Optimization**
- Online & Offline Results

Utility Loss

- Defined as the opposite of the Utility:

$$\ell_\sigma(p, y, v, c) := \int_0^{pv} (\tilde{c} - yv)\,\Pr(\tilde{c} \mid c, \sigma)\,d\tilde{c}.$$

- Non-convex; very hard.

Utility Loss and Log Loss (NLL)

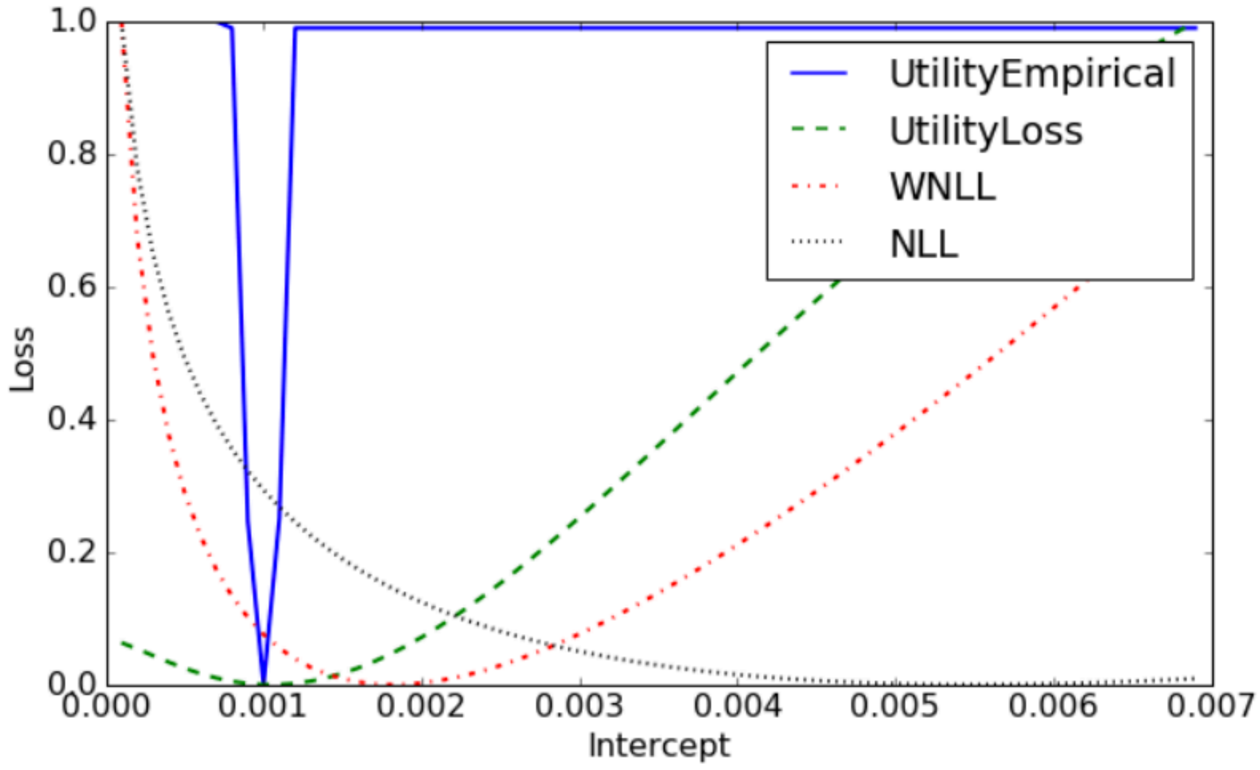
- We analyze the Utility loss when c is close to our bid pv.
- We assume conversion probabilities are small (p << 1).

	$y = 0$	$y = 1$
Log loss	$\frac{1}{1-p} \approx 1$	$-\frac{1}{p}$
Utility loss	v	$\frac{v(p-1)}{p} \approx -\frac{v}{p}$

=> The derivatives are approximately equal, up to a factor v

Toy Example

- Two advertisers with different CPAs (5 and 50) and CR (1% and 0.1%).



Method

- We propose to optimize for WNLL to improve our bidder's performance.
- We use L-BFGS for learning.

Impact on Regularization

- We propose the following heuristic to take weights into account:

$$\lambda_{WNLL} = \lambda_{NLL} \times \frac{\sum_i v_i}{N}$$

Agenda

- Context
- Online & Offline Metrics
- Utility Optimization
- **Online & Offline Results**

Offline Setup

- We use a public Criteo dataset for our experiments.

Offline Results – Weights

Weighting	ΔMSEW (negative is better)		ΔUtility$_{\beta=10}$ (positive is better)		ΔUtility$_{\beta=1000}$ (positive is better)	
	Train	Test	Train	Test	Train	Test
CPA	-50.45% ± 0.91	**-19.57% ± 0.65**	1.44% ± 0.02	**0.37% ± 0.04**	1.29% ± 0.03	0.18% ± 0.08
$CPA^{\frac{1}{2}}$	-36.84% ± 0.67	-14.57% ± 0.49	0.91% ± 0.01	0.32% ± 0.02	0.89% ± 0.03	**0.30% ± 0.04**
$CPA^{\frac{1}{4}}$	-24.54% ± 0.46	-9.26% ± 0.3	0.51% ± 0.01	0.18% ± 0.01	0.51% ± 0.02	0.19% ± 0.03

Offline Results – Lambda

Λ	$\Delta\text{Utility}_{\beta=1000}$
$\lambda_h - 40\%$	$0.34\% \pm 0.09$
$\lambda_h - 20\%$	$0.32\% \pm 0.05$
$\lambda_h - 10\%$	$0.29\% \pm 0.05$
λ_h	$\mathbf{0.30\% \pm 0.04}$
$\lambda_h + 10\%$	$0.33\% \pm 0.03$
$\lambda_h + 20\%$	$0.21\% \pm 0.03$

Offline Results – High/low CPA

Metric	Global	HighCPA(>10)	HighCPA(>10), LowSales(<30)
ΔMSEW	-14.57% ± 0.49	-15.26% ± 0.53	-22.36% ± 1.02
$\Delta\text{Utility}_{\beta=10}$	0.32% ± 0.02	0.78% ± 0.06	1.70% ± 0.16
$\Delta\text{Utility}_{\beta=1000}$	0.30% ± 0.04	0.78% ± 0.11	1.72% ± 0.21

Online Results

- The A/B test was done on more than 1 Billion ad displays, on world-wide traffic. Our change resulted in a +2% lift in ROI.
- We observed significant savings in display cost + an increase in sales performance for the advertisers, especially on the campaigns with high CPA and low number of sales.

Conclusion

- Weighted Log Loss allows to get closer to both offline and online metrics in the context of online advertising auctions.

Thanks!