Comments slides for Tuesday, Sept 29: Online ads

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University

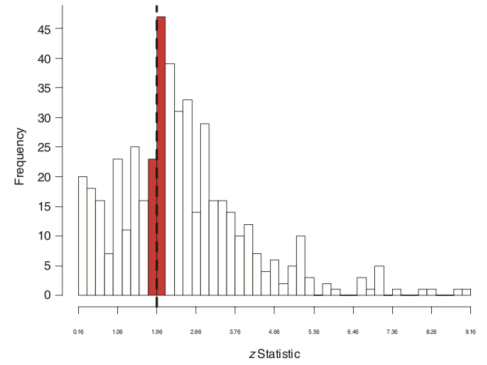
Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .

Figure 1
Histogram of z Statistics From the *American Sociological Review*, the *American Journal of Sociology*, and *The Sociological Quarterly* (Two-Tailed)



Are there fingerprints like this in the $\hat{y}$ world? Is it caused by researchers or journals?
Figure is from Gerber and Malhotra (2008)

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .
- In the case of ads, the predictor also partially controls the data generating process (compare to weather for example). How is that leveraged? How could it be?

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .
- In the case of ads, the predictor also partially controls the data generating process (compare to weather for example). How is that leveraged? How could it be?
- The observed ad click rates is not a good measure of the maximum possible click rate.

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .
- In the case of ads, the predictor also partially controls the data generating process (compare to weather for example). How is that leveraged? How could it be?
- The observed ad click rates is not a good measure of the maximum possible click rate.
- It is hard for anyone outside ad serving companies to know the accuracy of click predictions.

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .
- In the case of ads, the predictor also partially controls the data generating process (compare to weather for example). How is that leveraged? How could it be?
- The observed ad click rates is not a good measure of the maximum possible click rate.
- It is hard for anyone outside ad serving companies to know the accuracy of click predictions.
- There are still real computational issues.

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .
- In the case of ads, the predictor also partially controls the data generating process (compare to weather for example). How is that leveraged? How could it be?
- The observed ad click rates is not a good measure of the maximum possible click rate.
- It is hard for anyone outside ad serving companies to know the accuracy of click predictions.
- There are still real computational issues.
- They calculate AucLoss (1 - AUC), LogLoss, and SquaredError. Why not $? How should we balance between three metrics?

Observations/comments/questions/provocations:

- The readings about the complexity of measuring the effectiveness of ads is consistent with my experience. This is hard.
- Beware of estimates in fields where people are strongly incentivized to get a certain result. But sometimes this can leave fingerprints . . . .
- In the case of ads, the predictor also partially controls the data generating process (compare to weather for example). How is that leveraged? How could it be?
- The observed ad click rates is not a good measure of the maximum possible click rate.
- It is hard for anyone outside ad serving companies to know the accuracy of click predictions.
- There are still real computational issues.
- They calculate AucLoss (1 - AUC), LogLoss, and SquaredError. Why not $? How should we balance between three metrics?
- "Absolute metrics are misleading" They look at relative performance (% change relative to a benchmark model on exactly the same data). How many papers have we read with absolute measures (e.g., ImageNet)? When are absolute measures better? When relative? When do we only care about the order?

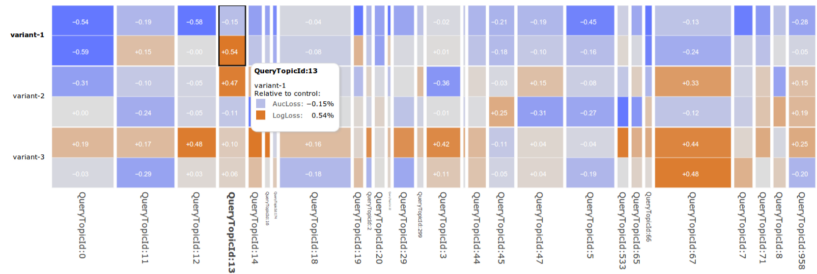


Figure 2: Screen shot of the high-dimensional analysis visualization. Here, three variants are compared with a control model, with results for AucLoss and LogLoss computed across a range of query topics. Column width reflects impression count. Detailed information pops up for a specific breakdown on mouse-over. The user interface allows for selection of multiple metrics and several possible breakdowns, including breakdowns by topic, country, match type, and page layouts. This allows fast scanning for anomalies and deep understanding of model performance. Best viewed in color.

Heterogeneity builds insight

Comments slides for Tuesday, Sept 29: Online ads

Matthew J. Salganik

COS 597E/SOC 555 Limits to prediction
Fall 2020, Princeton University