Response Prediction in Performance based Display Advertising

Krishna Prasad Chitrapura

pkrishna@yahoo-inc.com

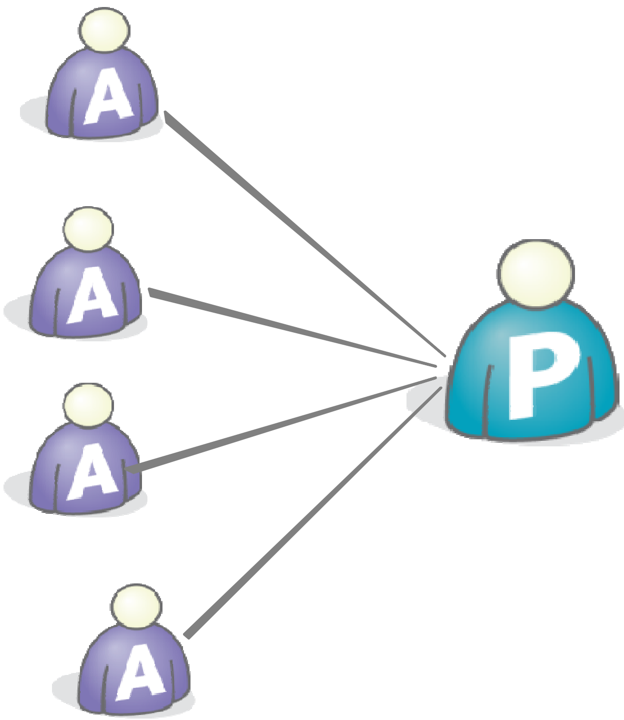
Yahoo! Labs, Bangalore

Overview

- History of display advertisement
  - From traditional print media to day.
- Why response prediction?
- Why is response prediction a hard problem?
  - What worked and what didn't
- Summary and pointers

History (Traditional media)

- Online display ad market mimicked the traditional media in early 90s. Publisher inventory was sold in bulk to advertisers
- Each Website was a publisher and involved lots of manual sales force.

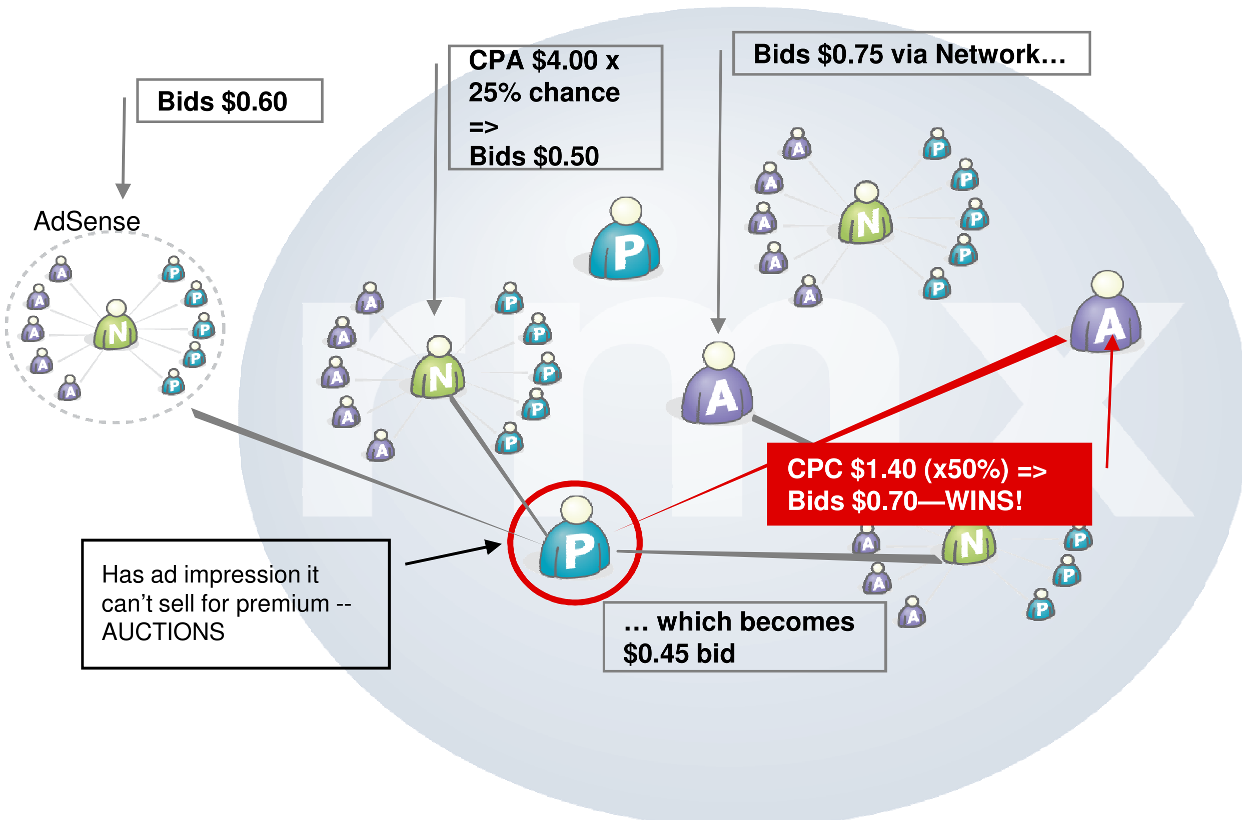


History (Network centric)

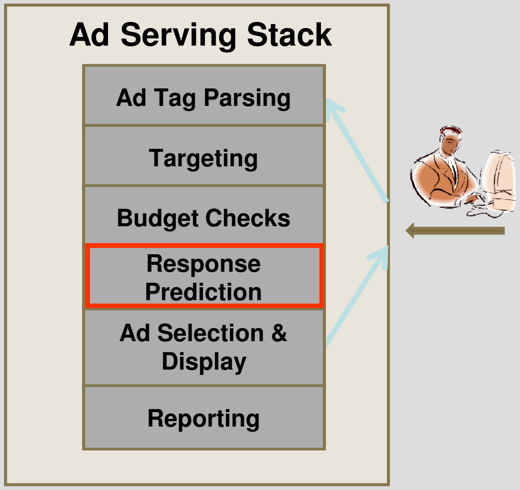
- This morphed into simple networks where aggregation of supply (eye balls) and demand (ads inventory) increased value. (Late 90s and early 2000).
- Each large publisher had a network – Yahoo! Aol, MSN, etc.
- Different pricing models evolved such as pay per click (CPC), pay per view (CPM), pay per user action (CPA).

History (Ad Exchange)

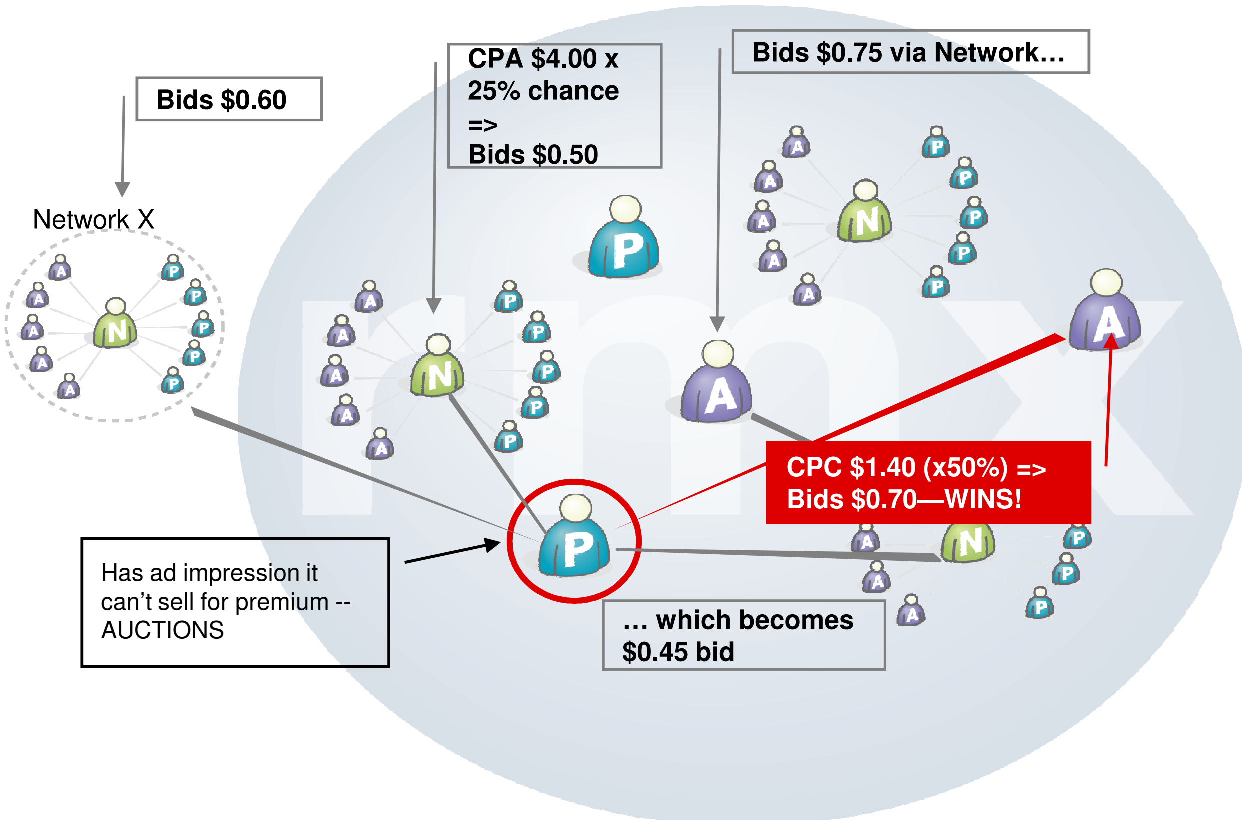
- Networks of networks resulted in better matching of supply to demand and had built-in incentives to share data about the inventory.



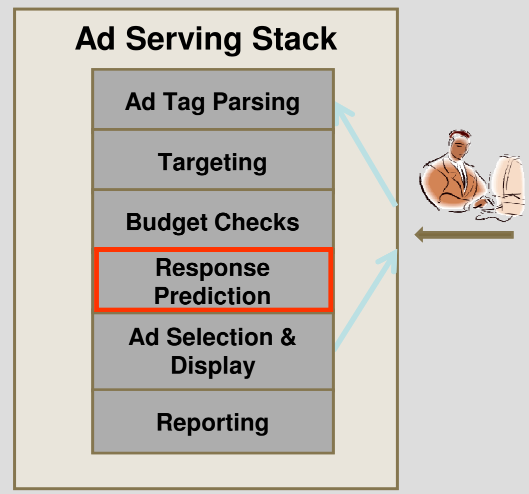
Need to compute the *probability* of click / conversion to help pick the *best ad*



History (Ad Exchange)

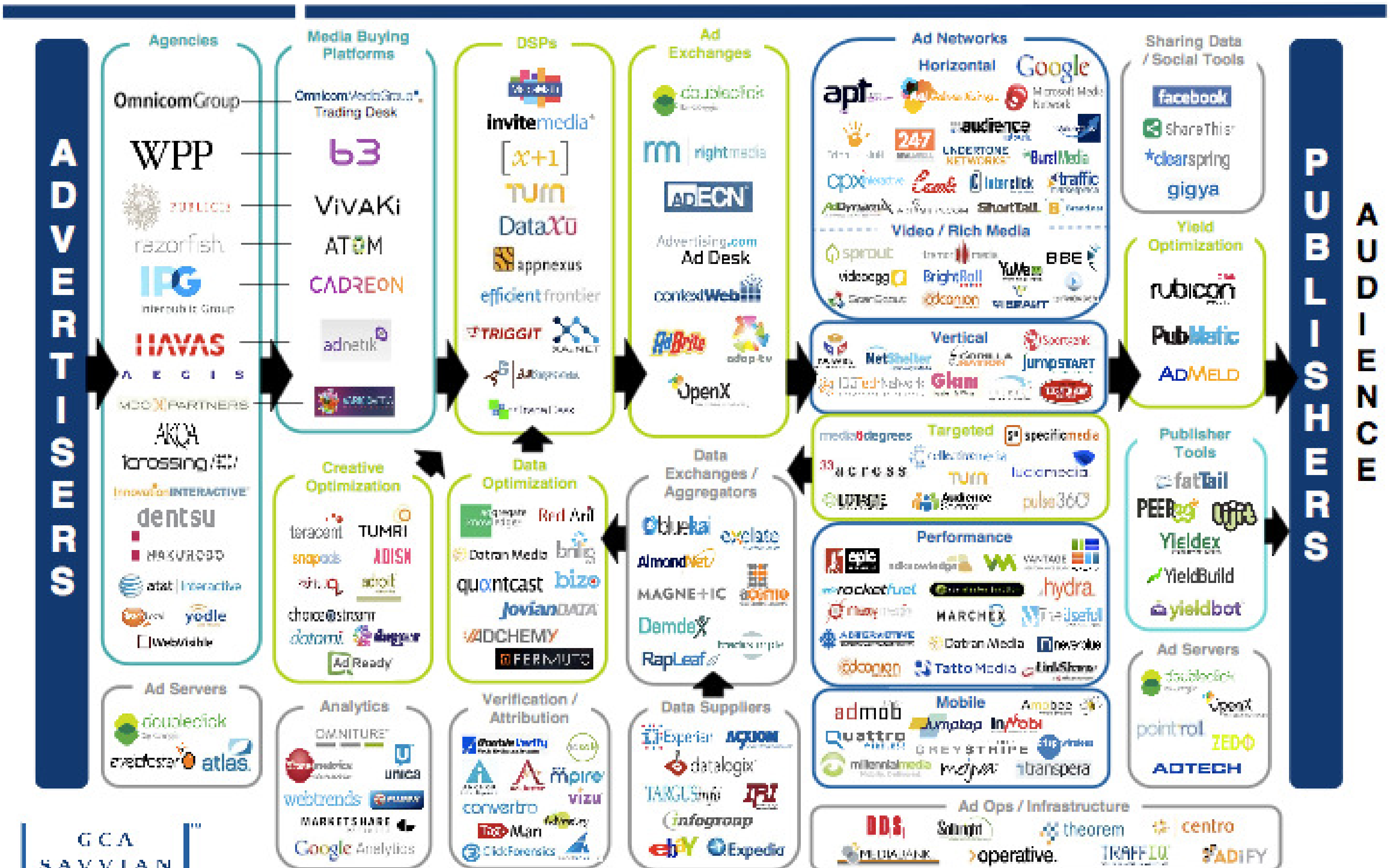


Need to compute the *probability* of click / conversion to help pick the *best ad*



Networks of networks resulted in better matching of supply to demand and had built-in incentives to share data about the inventory.

Situation Now (source: tkawaja@gcasavvian.com)



Please direct any inquiries or comments on the landscape to Terence Kawaja at tkawaja@gcasavvian.com or (212) 999-7082

GCA SAVVIAN

Problem(s):

- Extremely large dimensional space *O(B pages × 100 M user × few M ads)*
- Extreme sparsity – Click through rate and conversion rates are small.
- Data quality (frauds, missing value,…) and duplicity results in biased estimates.
- Is it a regression problem?
  - Not directly thanks to small rates.
- Is it a classification problem (c/ĉ)?
  - Not directly thanks to skew in class distribution

What was tried and tested?

- Linear regression/classification.
  - CTR estimates were not in [0,1].
  - Bounding did not help
- Logistic regression.
  - Sigmoid function does not separate the data well.
  - Moving to log scale helped although marginally.
- Latent Space models.
  - Regularized SVD with modifications to handle confidence helped.
- Smoothing the estimates on a hierarchy helped the most.

Response Prediction as Matrix completion

- Consider a *m X n matrix* of X of users demographics by advertisements.
- Each cell X_{ij} is the empirical click probability based on historical data. $X_{ij}=C_{ij}/V_{ij}$ for click and view matrices C,V.
- Problem 1: Smoothen X_{ij} for the observed cells.
    - Especially for cells very few views.
- Problem 2: Estimate X_{ij} for the unobserved cells.
    – If an ad has never been shown or has no clicks.

Matrix factorization

- Learn k latent features for the rows and columns via the non-convex optimization. (aka reguralized SVD)

$$\min_{U,V} \sum_{(i,j)\in O} (X_{ij} - U_i^T V_j)^2 + \frac{\lambda_U}{2}\|U\|_F^2 + \frac{\lambda_V}{2}\|V\|_F^2$$

- In addition, Fix U, optimize V. We need a ridge function to constrain the regression

$$\min_{V} \|X - f(UV^T)\|_O^2 + \frac{\lambda_V}{2}\|V\|_F^2 \qquad f : \mathbb{R} \to [0,1]$$

- Throw in confidence,

$$\min_{U,V} - \sum_{(i,j)\in O} C_{ij} \log \sigma(U_i^T V_j) + (V_{ij} - C_{ij}) \log(1 - \sigma(U_i^T V_j)) + \frac{\lambda_U}{2}\|U\|_F^2 + \frac{\lambda_V}{2}\|V\|_F^2$$

- Trust MLE for high view cells.

How does hierarchies help?

- The data can be arranged in multiple hierarchies
  - Ad hierarchy: Advertiser – campaign – ad_creative
  - Publisher : Publisher – site – sub-site-page
  - User : demographic hierarchy such as geo
- MLE at coarser levels of the tree have more confidence.
  - Smoothing/adding constraints based on coarser level bins reduces variance and bias.
- Couple of ideas:
  - Smoothen finer grained cells based on estimates in coarser bin.
  - Correct bias in finer grained bins based on aggregates in coarser bin.

More details

- Deepak Agarwal et al. "Estimating Rates of Rare Events with Multiple Hierarchies through Scalable Log-linear Models", KDD 2010.
- Adiya Menon et al. "Collaborative filtering using hierarchies for response prediction" under review
- Yahoo! Shares data to academic researchers under Webscope programme. See: http://webscope.sandbox.yahoo.com/