DISCUSSION ON "THE VALUE OF INFORMATION IN MOBILE AD TARGETING"

Sridhar Narayanan

Stanford University

Marketing Science-FTC Conference

16 September 2016

OVERVIEW OF THE PAPER

- Research Question
  - Ad networks have historical information on consumers, and can share data with advertisers at different levels
  - What is the value of this information to the network and to advertisers in predicting clicks by consumers
  - What kind of information is valuable?
    - Aggregated to the level of an ad, an ad within an app, at the impression level for each advertisers' impressions, impression level across advertisers?
  - To Whom?
    - Ad network, Big vs. small advertisers

OVERVIEW OF THE PAPER

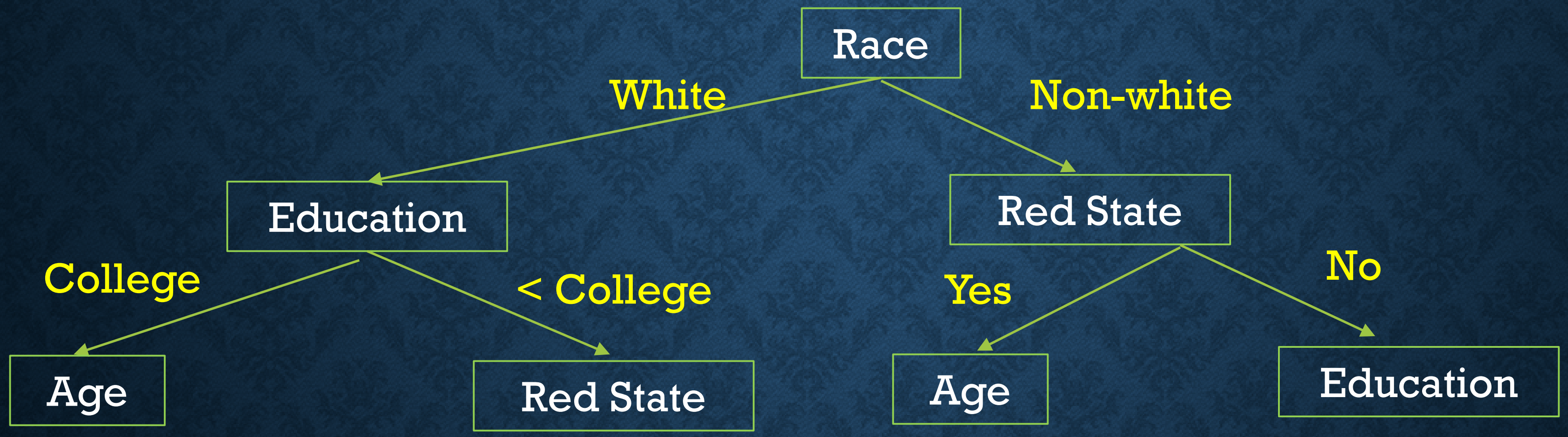
- Overall approach
  - Build a prediction model for predicting clicks by consumers
  - Use historical information to build a set of predictor variables
    - At various levels of aggregation over users, ads, apps, time
  - Compare different approaches on prediction accuracy
    - Standard econometric models - OLS, logit
    - Machine learning algorithms – multiple additive regression trees (MART)
  - Compare different information scenarios on prediction accuracy
  - And on click-through rates if the platform were to allow advertisers to target ads to specific impressions
    - For different sizes of advertisers

CLASSIFICATION AND REGRESSION TREES

- Problem – predict some outcome variable using a potentially large set of predictor variables
  - Linear or polynomial regressions assume that there is a *globally valid* relationship between predictors and outcome in the entire data space
  - Relationships may in reality differ across different subspaces of the data

- CART recursively partitions the data space based on a variable at a time
  - Such that outcomes are differentiated across the partitions and homogenous within
  - Looking forward and without revisiting prior partitions (greedy algorithm)
  - Does not guarantee a globally optimal solution (in fact it is not feasible as it is an NP-complete problem) but approximates it using a sequence of locally optimal solutions

EXAMPLE – PREFERENCE FOR PRESIDENTIAL CANDIDATE



BOOSTED DECISION TREES

- Prediction error consists of bias and variance
- Classification trees tend to have high bias although they have low variance
- Overfitting problem, high reliance on variables with multiple levels
- Boosted decision tree – reduces the bias by averaging across multiple decision trees
- MART – gradient descent boosting – the next decision tree is based on steepest descent in prediction error

MAIN RESULTS

- Ad-network's problem
  - MART does better than alternatives on relative information gain over the baseline (which assumes the average click through rate across all observations)
  - Using all information at app, ad and user-level leads to greatest gains
    - User-level information leads to more gains than app or ad level information (conditional on model used)
    - App-level information more useful than ad-level information (subject to same caveat)
    - User-id more useful than IP

MAIN RESULTS

- Information sharing problem
  - Highest gain in prediction when advertisers are provided impression-level data on their own ads
  - Lower gains when information across all advertisers is shared – softens competition
  - Bigger advertisers gain the most through targeting using historical data on their own impressions – they have more data with more variation

SOME COMMENTS

- Important problem
  - From the ad-network's perspective
  - Public policy implications as well
    - Privacy issues
    - Inefficiencies
- Good data
  - Rich, detailed data at the impression level
  - Auction mechanism induces some degree of randomization (more on that)
- Competent empirical work
- Interesting results confirming many of our intuitions on the issue

SUGGESTIONS

- Why not make the model comparisons more comprehensive? Why MART alone?
  - Authors refer to some prior empirical work establishing its superiority
  - But that is under specific conditions, and when averaging across metrics
  - Several other promising candidates, some alleviating some of the issues of MART
- Is this really value of information – more clarity in the phrasing and positioning would be useful
  - More clicks need not imply more value

SUGGESTIONS

- Do differences between large and small advertisers reflect a qualitative difference or a difference in the degree of randomness
  - Bigger vs. smaller advertisers differ in the degree of randomness in ad placement
  - Need to be careful in interpreting the resuls

TO CONCLUDE

- A nice paper that brings in machine learning tools into an important marketing and policy question

- The paper uses very good data, and applies it in a careful way

- With a more comprehensive analysis on the model comparison, and a more careful statement of the results, it provides a nice contribution to multiple literatures