When Structural Model Meets Big Data:
Examining Multi-Device Attribution for Native Ads Using TB-Sized Data

Quan Wang[1], Beibei Li[1], Pengyuan Wang[2], Jimmy Yang[2]
[1]Carnegie Mellon University
[2]Yahoo Labs

1. Introduction

Recently, researchers have adopted structural models to explain the behavioral cause and decision process of consumers. However, due to computational burdens, it is challenging to apply structural models to real-world datasets at very large scale, which to some extent has constrained the generalizability of such methodologies. Practitioners, on the other hand, have collected massive data but often turned to reduced-form models or “black-box” machine learning models for solutions which hardly have theoretical support. In this study, we aim to combine behavior-driven structural model with machine learning tools, and apply them to analyze real-world problem on a very large scale.

More specifically, we study the channel interdependence and multi-device native ads attribution at individual user level. We explore how the distribution of ads on multiple digital devices (i.e. tablet, smartphone and PC) work together to stimulate final conversion. This paper is among the first research studies to understand how tablet as an advertising channel interacts with other digital channels. To achieve our goal, we model the marginal impact of an advertisement impression based on the conversion funnel theory. We apply the Hidden Markov Model (HMM) to examine the attribution problem (e.g., Abhishek et al. 2012).

We validate our study on a unique dataset containing TB’s records that include billions of impressions for thousands of unique advertisers. Note that because digital conversions are rare events, it requires a large scale of individual-level data to acquire enough conversion events. To handle the sheer volume of the impression level data and the iterative nature of the estimation procedure, we develop a novel estimation algorithm that can distribute the data and computational burden in parallel on cloud computing infrastructure (Apache Spark and MapReduce). Our preliminary results show channel diversity stimulates disengaged consumers to transition to engaged stages. But it does not encourage the already engaged consumers to stay engaged. In addition, the channel diversity is more effective for the early impressions than for the late impressions.

2. Research Context and Data Description

This study focuses on an emerging type of advertisement: native ads. In contrast with traditional banner ads that fight with media content for space and attention, native ads match the visual design, format, and display location of the organic content. Our dataset contains TB's of ad impressions over two weeks' time period, including billions of impressions for thousands of unique advertisers (The description of the dataset are deliberately vague for confidential purpose). Each entry in our dataset comprises the device type, application type, campaign id, advertiser id, advertiser industry, whether it is clicked or not, whether it leads to a conversion or not, user id, user gender, user age, date and time.

We first provide the model free summary statistics about the conversion behavior. In particular, we compare the conversion rates for seven combinations of device(s) that consumers use to see the advertisement. The result using the data of an anonymous advertiser is reported in Figure 1. The heights of the conversion rates are different, suggesting advertising channel may play a role in consumers' propensity to purchase. When multiple advertising devices are used, the conversion rates are different than when a single device is used, suggesting that there can be interdependence in the advertising channels.

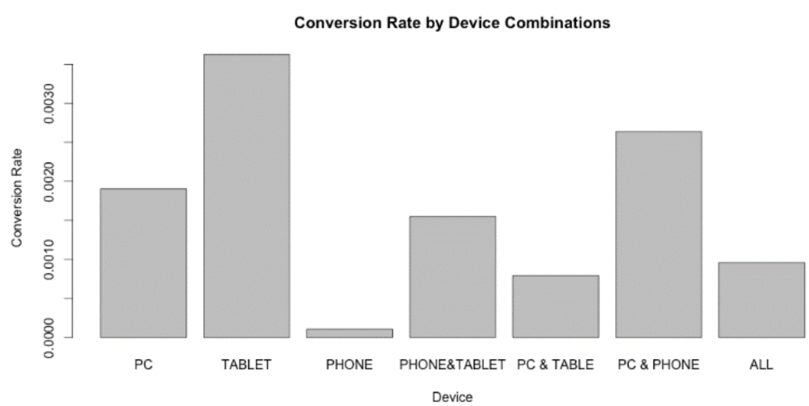

Figure 1. Conversion Rate by Device

3. Model Setup

Our structural model follows the conversion funnel framework. The advertisement fits in the framework by affecting when and how consumers move between stages. It is hypothesized that complementary advertising devices stimulate consumers to move to more engaged stages, whereas substitutive advertising devices stimulate movement to less engaged stages. In accordance with the conversion funnel, we construct a Hidden Markov Model (HMM) with three components: the observed conversion results for consumer i (denoted as C_i), the observed advertising exposures (denoted as a_i), and the corresponding unobserved stages (denoted as S_i).

In our model, there are three groups of model primitives about the transition matrix, conversion matrix, and initial distribution. We use the transition matrix Qit to represent the transition rule of hidden stages. The element qitjk in the jth row and kth column of Qit denotes the transition probability from state j at t to state k at t+1 for consumer i. The functional form of the transition probability is assumed as

$$q_{itjk} = p(S_{it} = s_k \mid S_{it-1} = s_j) = \frac{\exp(a'_{jk}\beta_{jk} + e_{itjk})}{\sum_{l=1}^{|S|}\exp(a'_{jk}\beta_{lk} + e_{itlk})}.$$

Consumer's probability to convert depends on the state in which she is currently staying. We model it as a state specific constant that is the same for all individual consumers. Finally we model the initial state membership as a function of individual demographics such as user age, gender, and an unobserved shock.

4. Estimation Strategy

When dealing with large scale dataset, it is usually not possible to use the same computational resources as that of a small dataset. This is due to two challenges in big data analysis. First, the sheer volume of our data makes standard implementation computationally unworkable. Second, the iterative nature of model estimation requires to reuse intermediate results across multiple iterations. To solve the big data challenges, we take the following novel estimation approach.

We first carefully choose the expectation-maximization (EM) algorithm (e.g. Rabiner 1989, Sahoo et al. 2012) to estimate our HMM model. Compared with alternative estimation algorithms such as Markov Chain Monte Carlo (MCMC) procedure or the original maximum likelihood estimation (MLE), the EM algorithm is much faster with linear time complexity.

Then we take advantage of nice properties of the EM algorithm so that the computation and storage can be conducted in parallel. The essence of our estimation procedure is to identify intermediate individual specific statistics that can be handled independently by mappers and reducers. The mapper is computationally intensive but the input data is divisible. The reducer is computationally easy but it uses all information of the entire dataset to update the model primitives.

Finally, the iterative nature of the EM algorithm makes Hadoop MapReduce, the most widely accepted big data technique, inefficient to solve our problem. Instead, it is necessary to leverage the state-of-the-art big data processing technique called Apache Spark. Spark was purposely designed to overcome the inefficient handling of iterations in MapReduce. It proposes a distributed memory abstraction that supports in-memory computation on large clusters. Using Apache Spark in our estimation, the original data is loaded only once. We implement the estimation procedure using Spark 1.3.0.

5. Preliminary Estimation Results and Model Applications

We apply the proposed model to a subsample with one representative advertiser. Due to the limited space, we only report the main results about channel interdependence.

The estimated transition parameters are reported in Table 1. From left to right, each column reports the estimated coefficients corresponding to a pair of transition stages. For instance, the column β_{dd} corresponds to the transition of staying in the disengaged stage. The result suggests the channel interdependence relationship may be affected by the current stage of the consumer. On one hand, the signs of the parameters -0.011, 0.341, -0.006 and 0.344 indicate that higher device diversity can significantly encourages the movement from disengaged stages to engaged stages. In other words, PC channel, smartphone channel, and tablet channel can be complementary for disengaged consumers. On the other hand, the sign of 0.007 and -0.266 indicate the opposite movement when the consumer is already in the engaged stage. Adding diversity of devices will reduce the engagement level, if consumers already have high propensity to buy. It suggests the three advertising channels can be substitute for engaged consumers.

	β_{dd}	β_{de}	β_{id}	β_{ie}	β_{ed}	β_{ee}
Number of unique devices	-0.011 (0.005)	0.341 (0.023)	-0.006 (0.003)	0.344 (0.066)	0.007 (0.011)	-0.266 (0.019)
Number of unique campaigns	-0.044 (0.009)	-0.365 (0.053)	-0.018 (0.021)	-0.295 (0.197)	-0.010 (0.011)	0.593 (0.675)
Number of clicks	-0.273 (0.023)	2.455 (0.643)	-0.275 (0.065)	2.357 (0.091)	-0.001 (0.006)	1.918 (0.721)
Time indicator	-0.014 (0.122)	-0.623 (0.832)	0.013 (0.048)	-0.588 (0.232)	0.017 (0.012)	-1.755 (1.328)
Intercept	-0.170 (0.025)	-1.378 (0.832)	-0.107 (0.034)	-1.321 (0.981)	0.008 (0.002)	-2.340 (1.731)

Table 1: Transition Matrix of a Typical User

The estimation results can be used to generate rich model applications. One example application is to recover the most likely path to conversion. This analysis provides a sequence of hidden stages that a consumer is most likely to have experienced. Moreover, aggregating the estimated hidden sequences for all consumers, we can plot how many consumers belong to each of the engagement levels over time (as in Figure 2). This figure shows that the proportion of engaged consumers increases for the first 4 or 5 advertising impressions. But the increasing trend does not hold any more if more advertising impressions arrive. Such non-linear time trend also hold for the less engaged levels.

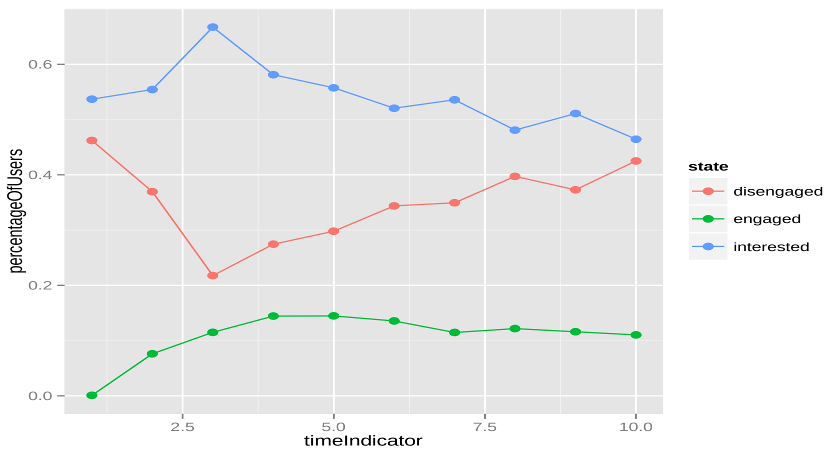

Figure 2. Time Series Plot of the Engagement Levels

6. Conclusion and Future Work

In this paper, we present a model that analyzes how consumers behave when they are exposed to advertising from multiple digital channels. A consumer moves through the purchase funnel in a stochastic manner that can be captured by a HMM model. Our estimation strategy allows us to handle massive dataset that is impossible with traditional estimation algorithms and computational resources. Our model can be used by business practitioners directly.

In the next step, we aim to leverage the rich dataset to verify the external validity of our conclusion. We will compare the estimation results for multiple advertisers from different industries. We plan to develop more model applications that address practitioners' business needs. Moreover, we are also interested in the value of big data versus the value of advanced models. We aim to contrast the model performance of two scenarios: one is the simple regression models using big data, the other is complicated structural econometric models using small data. This study has the potential to improve our understanding of consumer behavior in the hyper-connected digital world using structural models and big data.

References

Business Insider, 2015. Spending on Native Advertising is Soaring as Marketers and Digital Media Publishers Realize the Benefit. http://www.businessinsider.com/spending-on-native-ads-will-soar-as-publishers-and-advertisers-take-notice-2014-11

Lawrence Rabiner. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE.* 77(2). Pp. 257-286.

Michael Jordan and Tom Mitchell, 2015. Machine Learning: Trends, Perspectives, and Prospects. *Science.* 349(6245) pp: 255-260.

Microsoft, 2013. How Marketers Can Drive Cross-screen Engagement. http://advertising.microsoft.com/en/cl/1932/cross-screen-research-report

Sanjay Ghemawat, Howard Gobioff, Shun-Tak Leung. 2003. The Google File System. *Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles (SOSP).* New York, NY.

Sucharita Mulpuru, 2011. The Purchase Path of Online Buyers. http://www.canuckseo.com/wp-content/uploads/2012/10/Forrester_purchase_path_of_online_buyers.pdf

Tobias Ryden. 2008. EM versus Markov Chain Monte Carlo for Estimation of Hidden Markov Models: A Computational Perspective. *Bayesian Analysis.* 3(4). pp. 659-688.

Venture Beat 2015. Native Ads Are Replacing Banner Ads on Mobile – and Here's Why. http://venturebeat.com/2015/07/09/native-ads-are-replacing-banner-ads-on-mobile-and-heres-why/

Vibhanshu Abhishek, Peter Fader, Kartik Hosanagar, 2012. Media Exposure though the Funnel: A Model of Multi-Stage Attribution. *Working paper*. Carnegie Mellon University.

Zaharia, Matei, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauley, Michael J. Franklin, Scott Shenker, and Ion Stoica. 2012. Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing. *In Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation*, pp. 2-2. USENIX Association.