dstillery

Cross-Channel Measurement and ROI: Targeting Mobile App Usage to Increase Desktop Brand Engagement

Gilad Barash, Brian Dalessandro, Lauren Moores,
Claudia Perlich, Troy Raeder
Dstillery, New York, NY

Introduction

In today's world, the consumer experience is becoming more and more seamless across platforms, from laptop to tablet to smartphone. Marketers are challenged with harnessing the data collected from these different channels and using it to fully understand consumer behavior, in order to target users effectively across media and devices. As consumers spend more and more of their digital journey on mobile devices, the potential in mobile advertising is on the rise.

However, the mobile environment poses new challenges to marketers: how do we define and measure audiences in an effective way that allows marketers to selectively reach the right users? Today, a large proportion of mobile advertising is measured and optimized based on the click through rate (CTR) – defined as the number of clicks on an ad divided by the number of impressions served.

In this whitepaper we argue the fallacy of using mobile clicks for optimization or measurement. We provide evidence that the primary driver of mobile ad clicks is inventory hosting the ad, and that conditional on inventory, there is little correlation between CTR and the brand.

Additionally, we propose an alternative methodology for targeting mobile impressions for an advertising campaign. We show that a particular user's mobile app usage is correlated with the brands the user interacts with online. We present a methodology for measuring the correlation between mobile app usage and online brand engagement and show results for two brands that Dstillery works with. We show that mobile app usage can accurately predict online brand engagements and we argue that this methodology presents a better targeting mechanism than CTR optimization.

Our cross-device audience solution is achieved through the "crosswalk" - an innovative bridge that maps users between mobile and desktop platforms. This bridge allows us to follow consumers from one domain to another and create a combined physical and digital behavioral profile that allows us to find the right users for each marketer with high precision.

Avoid the Click

Measuring clicks and click-through rates (CTR) in any channel has long been an industry standard of evaluation. A lot of research has shown that clicks on online display ads are often not correlated with brand awareness or purchase intent [1]. We continue this line of research and examine clicks within advertising campaigns on mobile devices. We explore the question of whether mobile clicks provide any insight into a user's affinity toward certain brands? Our answer is, once again - not very likely.

Dstillery owns and operates a Mobile Demand Side Platform (DSP) with self-service functionality. Many large mobile advertising firms “plug-in” to the Dstillery mobile DSP and operate paid media campaigns for a large variety of brands. The Dstillery mobile DSP has access to thousands of mobile inventory sources across hundreds of millions of mobile devices via several mobile advertising exchanges (i.e., Mopub, AppNexus, Rubicon, etc.). This system processes billions of mobile events a day and tracks impressions and clicks across all campaigns. From this data stream we can assess exactly what drives clicks.

We were interested in the extent to which the branded message drives a user to click. To analyze this question we pulled data from two weeks of impression data across all active campaigns, representing several hundred brands. Our objective was to quantify the generalizability of different variables available at bid time for predicting a user click. We generated a set of models that utilized different groups of features, which included:

- Brand Creative (BC): Each campaign on average has multiple IAB standard formats. The creative id can be directly used as a variable that represents the brand. We encode each brand creative as a binary feature in a large yet sparse data matrix.
- Inventory (Inv): Each bid request includes a unique identifier of the mobile app or mobile web publisher that is originating the bid request. Each inventory source was encoded as a single binary feature.

We built predictive models using Logistic Regression trained with Stochastic Gradient Descent, an optimization algorithm that is generally the most robust in scenarios with large and sparse data [4]. Additionally, to avoid over-fitting our models we regularized our models using the L2 (or ridge) regression, as presented by [5]. We summarize the generalizability of our click-prediction model by using the Area Under the Receiver Operator Curve metric (AUC) [6]. The AUC is equivalent to the Mann Whitney U statistic and represents the probability that a positively labeled instance (i.e., a clicker) will have a higher predicted probability of clicking than a negatively labeled instance. Overall, it is a robust measure for the ability of a classifier to rank instances.

Figure 1 shows the in-sample and out-of-sample AUC for models built with three different combinations of the variables. The in-sample AUC represents two weeks of training data, and shows the extent to which the variables can explain the variance in observed click behavior. The out-of-sample AUC represents a week of campaign data that followed the training period, and shows the extent to which the variables can generalize and predict click behavior. For both samples, the inventory based variables are much more predictive than the brand creative based variables. What is more significant is that in the presence of the inventory variables, the branded creative information adds very little additional explanatory or predictive performance.

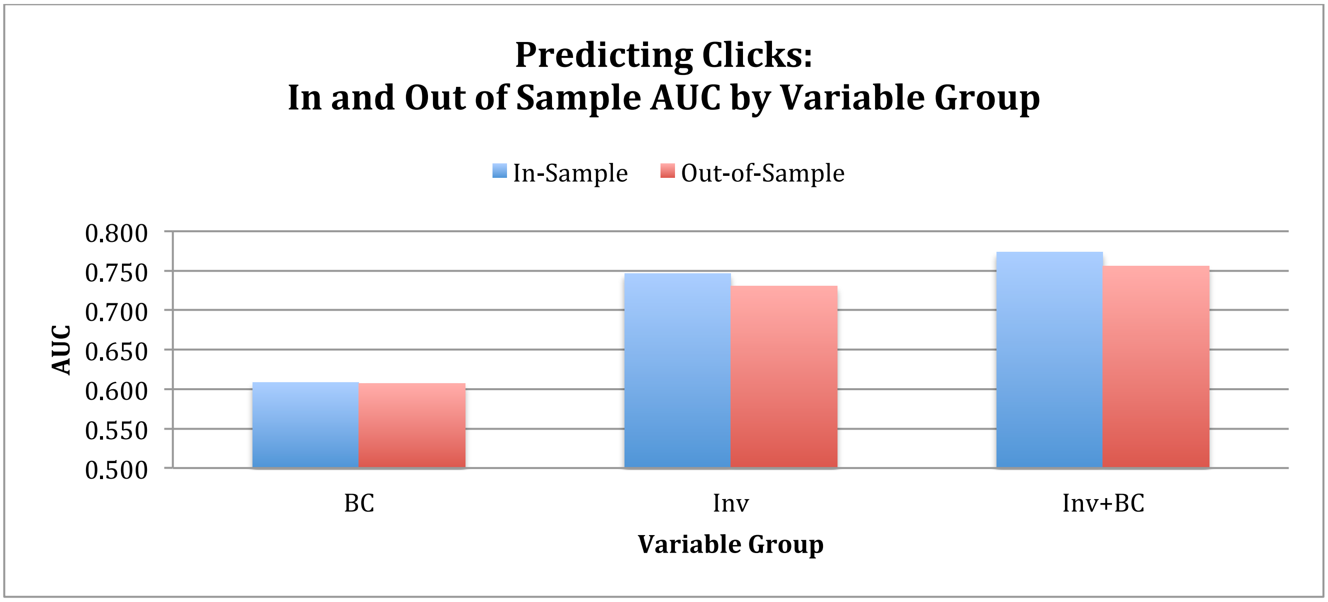

Figure 1: In and out-of-sample AUC results for click prediction models built with various sets of features. This experiment attempts to analyze the driver of click behavior and compares branded creative based variables with variables derived from mobile inventory. We see that inventory features generalize click behavior much better than branded creative features, and in the presence of inventory features, branded creative adds very little additional performance. The AUC ranges from 0.5 (completely random targeting) to 1.0 (a perfect classifier).

If the click was indicative of true interest in the brand or product, we would expect to see a stronger dependence between the CTR on and the specific brand being advertised. Instead we find that most of the observable dependence is on the inventory in question. When taking a qualitative look at some of the inventory, we find that many of the most clicked on mobile apps are flashlights and games that appeal to children. Figure 2 shows the IAB categories of the top 10 most clicked apps in the Dstillery DSP (instead of showing absolute CTR we index each app's CTR against the DSP average).

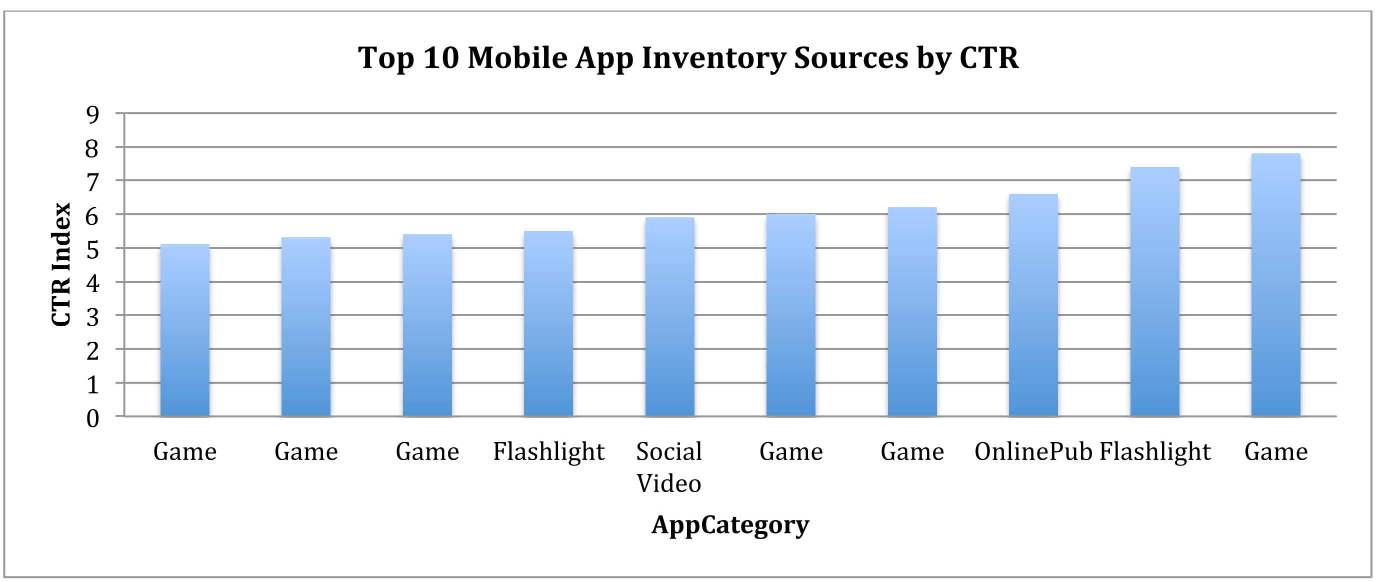

Figure 2: The IAB App category of the top 10 most clicked-on apps in the Dstillery DSP.

Our analysis shows that predicting and optimizing CTR is highly achievable in our mobile DSP. With a predictive model that selectively filters bid requests purely on inventory, one can achieve lifts of 5 to 10 times better than random targeting. Figure 2 shows a few examples though of the types of inventory a click optimizing system might select. Two of the top ten ranked apps by CTR are flashlight apps, implying that users literally clicking on ads in the dark. Several of the others are games that are targeted to children. This begs the question on whether or not many mobile clicks are intentional, and whether or not the clickers are even adults. As it stands there is no standard for correlating mobile clicks with more relevant brand engagements (such as purchases, store or site visits, etc.). Our analysis suggests that more research is needed to establish the viability of the mobile click.

With evidence that clicks are not good indicators of consumers' brand awareness or purchase intent, we investigate other mobile data points that might provide us the needed signal. Specifically, we need a targeting mechanism that can identify an audience that will be more likely have some positive engagement with a brand in question. Another such data point is mobile app usage, which gives us a means to segment the mobile user base in ways that are more aligned with specific brands.

Mobile App Usage Predicts Brand Engagement

Mobile apps are ubiquitous on our mobile devices. For any of our digital needs, the saying goes, "there is an app for that". Both our app and brand preferences are a reflection of our persona and tastes. Hence we hypothesize that there should be a correlation between brands that consumers engage with and the apps they use on their mobile devices.

Given the above-mentioned weaknesses of CTR as both a measurement metric and an optimization device, we seek an alternative method for selecting an audience that is more aligned with the brand. For that we look at mobile app usage as a starting point. To build a branded audience with mobile app data, we have developed a method for correlating app usage directly with brand engagement. For our study, we use interactions with a brand's owned website, such as site visits and/or purchases, as our signal for brand engagement. Using a technology that bridges mobile devices with desktop cookies, we can measure how brand engagement varies with app usage. With this set up we can then build a targeting mechanism that is more aligned with the brand than targeting based on CTR optimization.

Methodology

Our method starts with sampling users that have been observed on a particular mobile app. We then match the mobile device to a desktop cookie using a proprietary cross-device matching algorithm. With that bridge established, we can observe conditional probabilities of brand engagement. For each app, we define a simple metric that is scalable for large data environments, is well supported by query language implementation, and is easily interpretable. In this section we will present the methodology in more detail and define our validation procedure.

Data:

25 marketers were chosen as a basis for this experiment. This experiment combines data from two sources:

1. Dataset $\mathcal{D}1$ – A week's worth of anonymized mobile in-app bid-requests observed in online advertising exchanges, detailing the app from which the bid request originated.[1] From this data we sample users that have engaged with a particular app.
2. Dataset $\mathcal{D}2$ – A week's worth of anonymized, cookie based brand engagements. The brand engagements are observed via standard retargeting pixels that marketers implement as part of a standard Dstillery advertising campaign. We use this data as the outcome of interest and will correlate app usage to it.

Crosswalk:
To connect these disparate data sources, we create a device-to-device mapping using a "crosswalk" – an analytical bridge that connects mobile device app transactions to unique consumer transactions on desktop. This bridge uses probabilistic device matching to link users across devices and platforms based on a weighted random walk algorithm to determine link similarity [3].

The output of this process is a match between cookie and mobile phone, and for each matched pair we have the mobile apps on which we have observed the mobile device as well as the brand interactions associated with the cookie. We generate two such datasets using disjoint weeks. We use the first week as a training set and the second week for validation.

Correlation Metric
With the data we have sampled we can directly measure how brand engagement is conditionally related to mobile app usage. We propose that that if we can establish correlation between mobile app usage and desktop brand engagement then we can use app usage alone as a mechanism for targeting an audience.

We first design a metric – Brand App Index (BAI) - that measures correlation between app usage and brand engagement.

This is defined as:

$$BAI = \log\left(\frac{P(brand\ engagement\ |mobile\ app)}{P(brand\ engagement)}\right)$$

The Brand App Index (BAI) measures how much more (or less) likely the user of a specific app is to take brand action relative to the general population. This metric has several properties that are desirable as a targeting metric: 1). It is similar to the parameters learned in logistic regression and naïve bayes [7], and thus produces results similar to either classifier in a univariate case, 2). It is directly a function of the lift one should expect by targeting a particular app in a mobile bid stream, 3). It is easily computable using standard query language (in our system we use Hive, which as a relational database system that works over a Map-Reduce framework). Additionally, we use the log transformation to achieve a metric that is symmetric around zero and not overly skewed in the positive direction.

The BAI involves to probability estimations. In the numerator, we have the probability of brand engagement conditional on the user being in the set that has been observed on a particular app. We normalize this with the unconditional probability of brand engagement across the entire population. The BAI is calculated for each Brand and Mobile App pair in the system. The intended usage is to be able to rank mobile apps for each brand so that our mobile DSP can better optimize the allocation of impressions for each campaign – in other words, by targeting the apps with the highest BAI, we expect to find users that are more likely to be receptive to engaging with the brand.

Validation

The BAI can be positioned as a discovery and analytics tool, but mostly its intended use is as a predictive tool. We have thus developed two validation techniques to make sure our estimates generalize to an out-of-sample holdout set:

1. Directional Accuracy - the trend, or consistency, of over-indexing (or lift) of apps per marketer over time. To trust that the BAI is indeed a valid signal and not dominated by variance, this test measures the accuracy to which a particular app is a positive or negative predictor of brand engagement.
2. Rank Consistency - the correlation of the magnitude of over/under-indexing of apps per marketer. Ultimately, we wish to use the BAI as a ranking mechanism. This test measures the consistency in the magnitude of the BAI over multiple time-periods. We use the Pearson Correlation of the BAI across two time periods as our measure of rank consistency.

Predictive Performance

Figure 3 shows the BAI for a marketer over a set of mobiles apps for two independent time periods for a large Telco marketer. Each point represents a single app's BAI for the Telco marketer across two time periods.

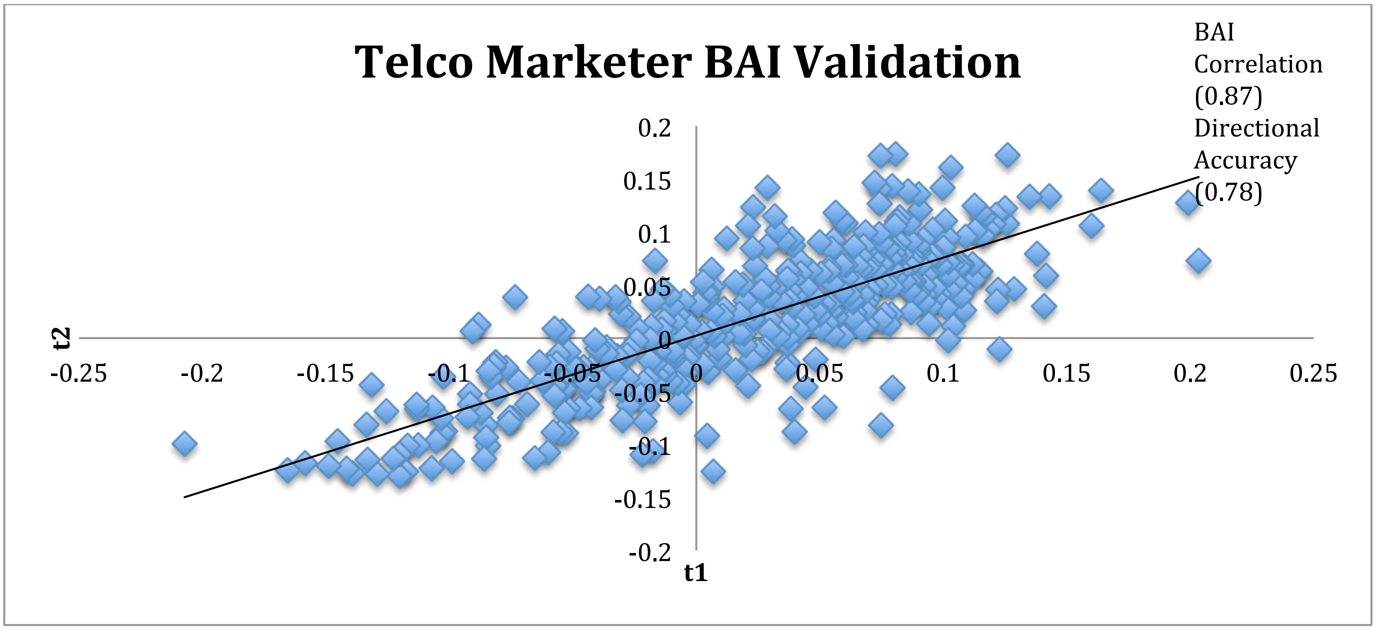

Figure 3: Correlation of high site visit propensity over two periods for a Telco marketer. The BAI correlation of 87% is the Pearson correlation between BAI across successive time periods. The directional accuracy of 78% is the likelihood of sign (BAI) to be the same over both periods.

We see a very strong consistency between t1 and t2. The directional accuracy of 78% means that an app that over or under-indexed in period t1 has a 78% probability of pointing in the same direction in period t2. We can see visually that directional accuracy is lower as the BAI approaches zero. In realistic targeting settings we would mostly only target apps with a higher BAI (such as BAI>0.05). In these cases the directional accuracy is better than 90%.

The BAI correlation in is 87% between successive time periods, suggesting that for this marketer our estimates are robust and will generalize across time. We thus concluded that this model is suitable for predictive purposes. This validation technique is akin to showing the predicted probability of brand engagement (from t1) is relatively close to the realized probability (from t2), and therefore it is possible to use mobile app to predict brand engagement. It is interesting to note that the list of predictive apps changes by marketer, and therefore can lead to specific and personalized strategic targeting.

In some instances, results showed correlations and consistencies that were less strong, but still provide a list of over-indexing apps that can be targeted by the marketer. Figure 4 shows results for a financial services company. Here we see the overall directional accuracy is 37% and the BAI correlation is 67%. Overall the results are less strong, but we can still see that in the more positive range of BAI the correlation and directional accuracy are better than the average.

dstillery

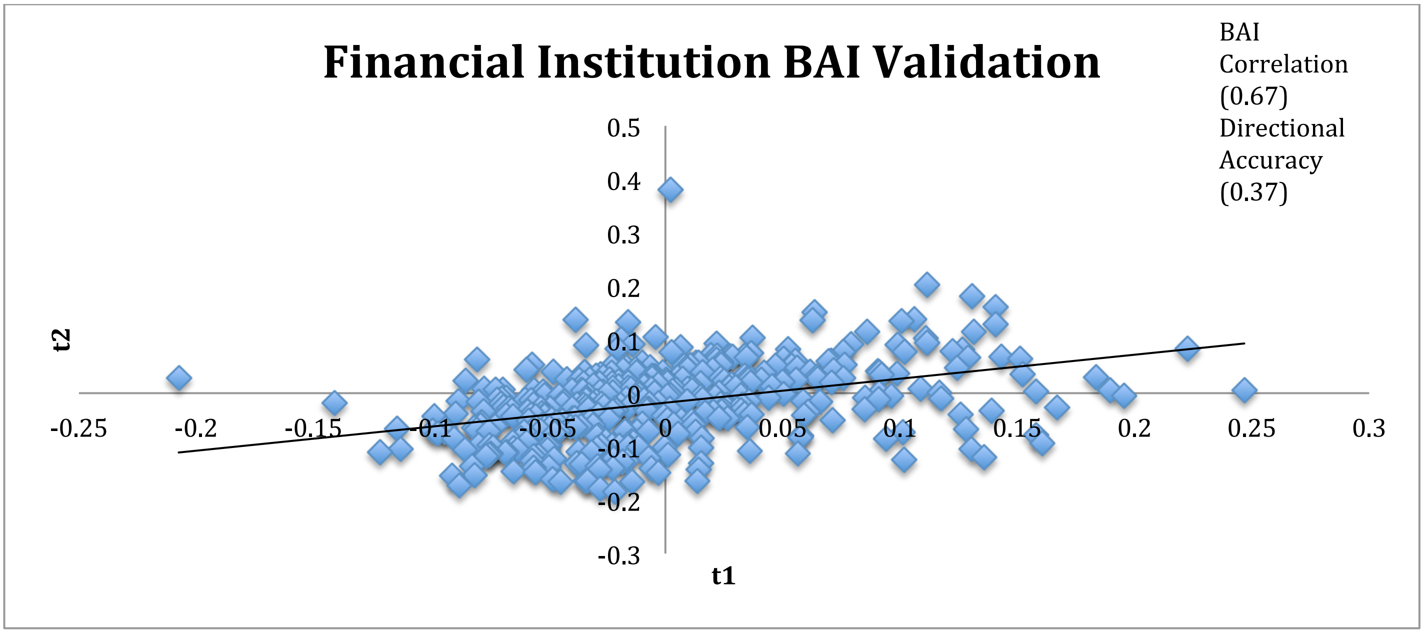

Figure 4: Correlation of high site visit propensity over two periods for a Financial marketer. The correlation of 37% is the Pearson correlation between BAI across successive time periods. The consistency score of 67% is the likelihood of sign (BAI) to be the same over both periods.

Discovery

An easy approach to developing an app targeting list would be to identify apps that are contextually, or categorically related to the marketing in question. Advertising based on matching ad creative with contextually similar inventory has long been a practice across multiple channels [8]. A complement to contextual advertising is behavioral targeting, which aims to match an ad creative to a particularly relevant audience, regardless of the context of the inventory [9]. Our methodology here can be regarded as a form of behavioral targeting.

Part of the process of building a behavioral targeting system is discovering inventory sources that are not contextually related to the advertiser in question, but nonetheless is relevant to the advertiser's audience.
Table 1 shows a sample of the apps that exhibit high propensity for the Telco marketer in question, along with the BAI for each of the time periods. By targeting the devices that show activity on these apps, we can optimize the audience for this marketer and drive performance. A benefit of the methodology we present here is that these apps would likely be overlooked in a contextually based targeting system. Nonetheless, a marketer could achieve lifts in brand engagement of 4-7x based on the estimates of BAI we have observed for these apps.

App Name	T1 BAI	T2 BAI
Gin Rummy Free	1.36	1.46
NOAA Weather Radar	2.07	1.55
WWF iphone	1.41	1.46
TVGuide mobile	1.26	1.42
iOS: Magic Puzzles	1.61	1.68
Talking Tom & Ben News – Android	1.82	0.97

Table 1: Sample apps that over-index for the marketer

Identifying mobile app usage that correlates to brand affinity also allows us to analyze usage across app categories and create an App Category Index. This index enables us to see which types of apps tend to over-index (or under-index, for that matter) compared to the general population. These may not be intuitively related to the marketer's industry vertical, since they emerge from the usage data that we analyze.

Figure 5 details such an index for a Financial marketer. It details iTunes app categories and compares usage of apps in that category between the marketer's brand-engagers vs. the general population. We can see that apps such as finance (understandable) and weather and medical tend to over-index (i.e. a larger percentage of this marketers users use financial and weather apps than in the general population). These users tend to use reference and book apps less than the general population, and therefore these app categories tend to under-index for the same marketer.

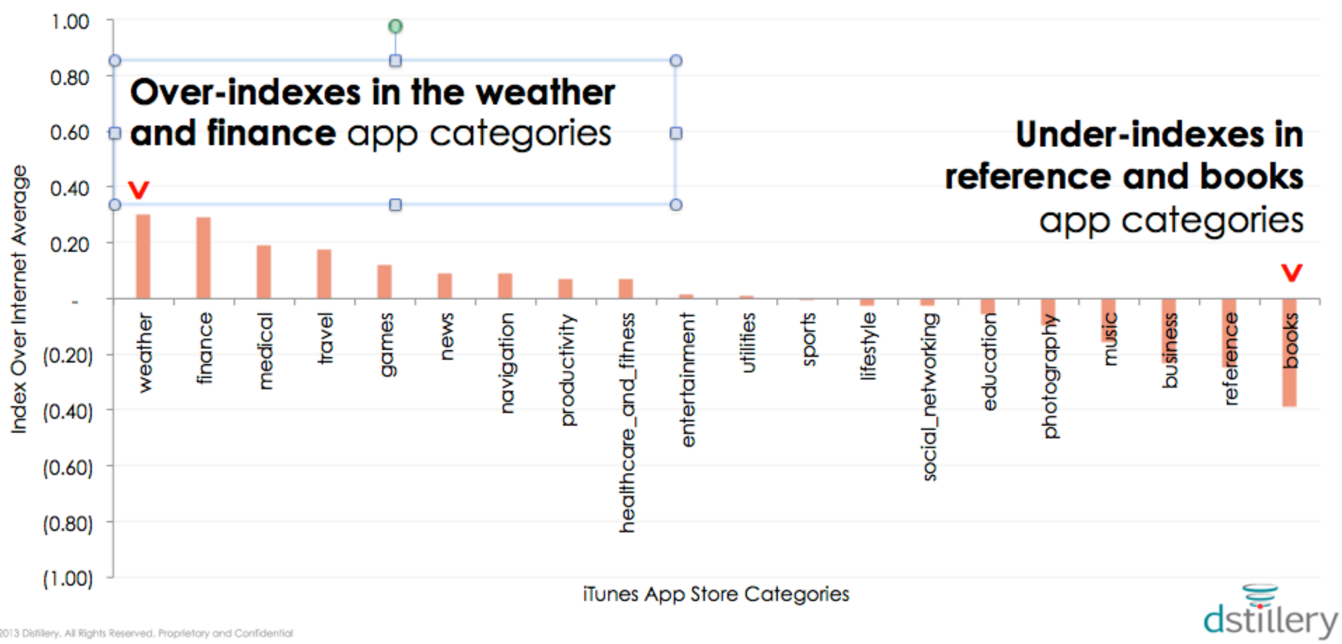

Figure 3: App category index for a financial marketer

Conclusion

Marketers looking to expand into mobile advertising or increase their current mobile advertising spend levels want to ensure the best ROI by reaching the right audience. In this study, we show that we can identify the right audience by using cross-channel, cross-device data to build mobile app profiles that are highly correlated with the user's brand preferences. By adopting this targeting and measurement strategy, we avoid using a substandard proxy for engagement - the click – and use instead a source of signal that is directly related to the advertiser's goal of driving more brand engagement. We conclude based on our BAI model that app usage is indicative of brand preferences and that we can effectively predict future brand engagements by targeting users of apps that have a high BAI.

References

[1] – Dalessandro, Brian, et al. "Evaluating and optimizing online advertising: Forget the click, but there are good proxies." (2012).

[2] - "The Click Remains Irrelevant: 'Natural Born Clickers' Return" was presented by comScore Chairman Gian Fulgoni on September 14, 2009, at the iMedia Summit in Coronado, CA. http://www.comscore.com/Insights/Presentations_and_Whitepapers/2009/The_Click_Remains_Irrelevant

[3] Provost, Foster, David Martens, and Alan Murray. "Geo-Social Network Targeting for Privacy-friendly Mobile Advertising." CeDER-11-06a

[4] Bottou, Léon, and Olivier Bousquet. "The Tradeoffs of Large Scale Learning." *NIPS*. Vol. 4. 2007.

[5] Hastie, Trevor, et al. *The elements of statistical learning*. Vol. 2. No. 1. New York: Springer, 2009.

[6] Provost, Foster, and Tom Fawcett. "Robust classification for imprecise environments." *Machine Learning* 42.3 (2001): 203-231.

[7] Ng, Andrew Y., and Michael I. Jordan. "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes." *Advances in neural information processing systems* 2 (2002): 841-848.

[8] Broder, Andrei, et al. "A semantic approach to contextual advertising." *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2007.

[9] Perlich, C., et al. "Machine learning for targeted display advertising: Transfer learning in action." *Machine Learning* (2013): 1-25.