July 19, 2017

Increased Efficiency In Ad Modeling Through Adveritser Analytic Power-User Analysis

Alex Jacobson

INCREASED EFFICIENCY IN MODELING THROUGH ANALYTIC POWER-USER ANALYSIS

In online content placement systems, content items such as advertisements can be displayed within information resources (e.g., webpages) to increase the chance that users will purchase, or take other actions with respect to, products or services associated with the content items. These systems can be highly performance driven, relying heavily on the accurate measurement of defined goals. Publishers, content distribution networks, content and application providers and other entities track user milestones following content item interaction (e.g. impression, click, view) to look for successful conversions (i.e. taking the desired action, for example, a purchase of a product/service or provision of desired information) and attribute them back to the last content item shown. Analysis of these conversions can come from analytic information resources or databases. In general, these reporting suites rely on content providers to express their goals through various rules or criteria. For example, some criteria may include page visits, user sign-ups, or purchases, to name a few.

One challenge with the above system is that setting these goals results in high variability across verticals and even across content providers within the same vertical. Although some content providers articulate helpful and useful criteria to measure, many fall short, especially when conversion history is not attributable to a user cookie. For example, some content providers set goals that are too easy (e.g. user arrived at landing page), and others set goals that are too difficult (e.g. user purchased a diamond ring on the website).

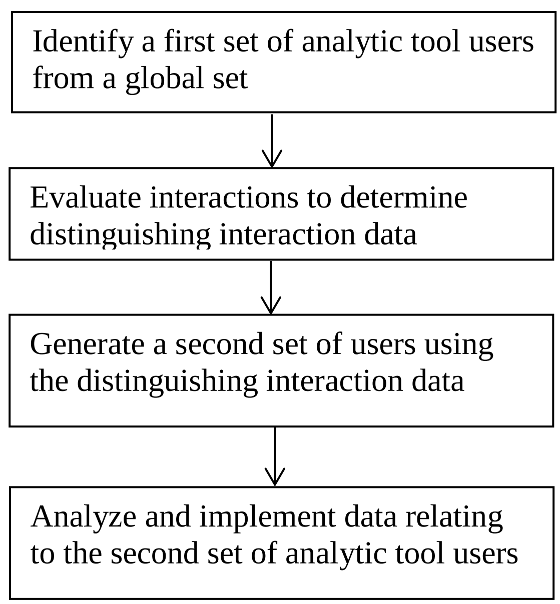
As a result, several negative effects can arise. The problems are compounded due to the extent that conversions are used for modeling bidding, pricing, and quality. Not only does this deficiency in measurement hurt the content provider, but also it can cause a larger content item network to operate in a sub-optimal standard. This may negatively affect the performance of content providers as well as decrease revenue for other entities and the network.

Conventional methods to address the foregoing generally only provide marginal improvements. For example, some improvements can be realized by removing sufficiently high and low data points, thereby decreasing variability. This approach may only provide minimal benefits, and does not facilitate future improvements. Other analytical studies can be time-intensive and may not uncover subtle effects.

In order to overcome this issue, increased efficiency in modeling through analytic power user analysis is introduced herein. Implementations involve the construction and classification of content providers through study of analytics tool usage data to identify "power users" across the network who are particularly proficient with the analytics tool. An analytics tool is used by content providers that allows interactions with the tool to be stored in an analytics database for subsequent evaluation. Examples of interactions include selection of features or settings, setup and tracking of content items, used or unused options, report usage, etc. Information or actions relating to interactions can also be monitored (e.g. the frequency of an interaction). A subset of training data can be identified manually to find conversions that are well tracked and correlate with other content item studies. Using this training data, the contents of the analytics tool usage can be an input for Principal Component Analysis (PCA analysis) to identify previously unrealized settings that identify content providers that are especially good at designing and deploying goal measurement. Using this cluster of content providers, measurement accuracy relating to the performance of content items (e.g. for modeling, pricing, and other auction study) can be improved.

The foregoing method more closely correlates the milestone setup process and reporting tool usage to identify content providers who are particularly good at tracking goals. Studying the logs in this manner can reveal unidentified correlating effects, such as what reports are studied, the frequency of checking in, deep advanced features that are deployed, combination(s) resulting in good tracking behavior, etc. These learnings can help teach best practices to other content providers or otherwise automate analytic goal creation.

One implementation of the increased efficiency in modeling through analytic power-user analysis is illustrated in the flowchart below:



An offline analysis tool may be used to facilitate the foregoing implementation, in some implementations. The offline analysis tool is configured to identify a first set of analytic tool users from a global set of analytic tool users. The global set of analytic tool users generally includes content providers (persons or entities) who interact with a particular analytics tool. Interactions of each user may be stored in a centralized analytics database, which can be configured to communicate with the offline analysis module via a network.

Interactions can include any action or behavior relating to designing and deploying goal measurement. For example, some interactions include configuring settings, swiping, clicking, dragging, dropping, etc. Other examples of interactions can include selected features or settings, default settings, setup and tracking of content items, used or unused options, report usage, etc. Information or actions relating to interactions can also be stored in the analytics database, such as the time and day of an interaction, the frequency of an interaction, etc.

The first set of users generally includes users from the global set who are identified as particularly proficient in setting up and/or tracking content items (e.g. goals) relative to the global set. The first set may be identified by examining one or more interactions stored in the analytics database. For example, the first set may be identified based on a total amount of time the user uses the analytics tool or a level of experience associated with the user. Implementations can additionally or alternatively use any combination of parameters or factors indicative of proficiency. For example, a factor indicative of proficiency may relate to investment associated with the analytics tool. In some implementations, the first set can be identified based on a recommendation or identification from one or more persons, for example due to existing relationships with the user.

The offline analysis module is further configured to evaluate interactions of the first set of analytic tool users to determine distinguishing interaction data. The determined distinguishing interaction data generally relate to interactions or information associated with the interactions that distinguish the first set of users from the global set of users. In some implementations, the offline analysis module retrieves data from the analytics database associated with each analytic tool user. The retrieved data includes interactions and/or information associated with the interactions of each user. The offline analysis module can evaluate the retrieved data to determine distinguishing interaction data, which includes interactions and/or associated information that distinguish the first set of users from the global set. Any suitable method can be used to determine the distinguishing interaction data, such component reduction methods, principal component analysis, etc. In some implementations, distinguishing interactions data are determined based on a comparison of an average value of an interaction associated with the first set of users to an average value of the interaction associated with the global set of users. For example, one interaction can be usage time associated with the analytics tool. An average usage time of the first set of users can be calculated and an average usage time of the global set can be calculated. The calculated average times can be compared to determine whether usage time is a sufficiently distinguishing interaction. Any number of interactions and/or information associated with an interaction can be evaluated.

The offline analysis module is further configured to generate a second set of analytic tool users using the distinguishing interaction data. The second set of analytic tool users is a subset of the global set and generally corresponds to a larger number of proficient users relative to the first set. The offline analysis module is generally configured to identify the second set of analytic tool users based on the distinguishing interaction data. The offline analysis module can evaluate interactions and associated information of each user of the global set to determine whether each user's interactions correspond to the distinguishing interaction data, and thereby generate the second set of users. In this regard, the second set includes users in addition to the first set that can also be characterized as proficient users based on similar interaction behaviors and/or patterns. For example, the distinguishing interaction data can relate to tool usage time greater than two hours per week, usage of a particular analytics report every week, and configuration of a setting as a default setting. The offline analysis module can retrieve data from the analytics database relating to interactions of each user in the global set to determine whether each user's interactions correspond to the distinguishing interaction data.

The offline analysis module is further configured to analyze and implement performance data relating to the second set of analytic tool users. The performance data generally relates to data (e.g. interactions and/or associated information) corresponding to the second set of users. In some implementations, the offline analysis module is configured to retrieve stored performance data from the analytics database relating to each user in the second set of users. In some implementations, a distinguishing interaction and/or associated information can be used for improving performance of one or more users (e.g. suggesting a particular tool setting, including a particular dataset in a report). In some implementations, modeling across one or more industries can be improved by using the performance data of the second set of users rather than, for example, data relating to the global set of users. For example, improved modeling precision (e.g. quality measurements, targeting accuracy, etc.) may allow identification of smaller changes due to smaller confidence intervals.

Abstract

Increased efficiency in modeling through analytic power user analysis is introduced in this publication. The analytic power user analysis includes identifying a first set of users from a global set of users. The first set of users is generally characterized as proficient, relative to the global set, in setting up and/or tracking a content item. Interactions of the first set of users are analyzed to determine interactions or information associated with interactions that distinguish the first set of users from the global set. A second set of users are identified based on the distinguishing interactions or associated information. Performance data associated with the second set of users is used for analysis and modeling improvement.