December 2019

MODELING STATISTICALLY SIGNIFICANT PROPENSITY CONTROLS

Brian Mulford

Yueheng Du

Colby Ranger

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

MODELING STATISTICALLY SIGNIFICANT PROPENSITY CONTROLS

A content item service may allow users to upload content items (e.g., videos, songs, audiobooks, images, documents, etc.) on the content item service. Such content items may be streamed or otherwise provided or rendered to various users. Content items can include, and are not limited to, digital video, digital movies, digital photos, digital music, website content, social media updates, electronic books (ebooks), electronic magazines, digital newspapers, digital audio books, electronic journals, web blogs, real simple syndication (RSS) feeds, electronic comic books, software applications, etc. A content item may also be referred to as a media item. For purposes of the discussions herein, a content item may refer to a video.

The content item service may allow brand advertisers to run advertisement campaigns utilizing the various content items streamed or otherwise provided or rendered by the content item service. The content item service may measure attitudinal response to the advertisements of the advertisement campaigns run on the content item service. For example, the content item service may measure impact on brand recall after being exposed to branded content. This allows the content item service to help advertisers understand advertisement campaign performance and grow revenue.

To measure attitudinal responses of users, the content item service may utilize studies, such as a survey lift analysis or survey lift study, to determine the attitudinal responses of users of the content item service to the advertisements. Such studies may also be referred to as advertisement campaign studies herein. Designing such studies may involve identifying appropriate groups of users for the studies, including generating an exposed group and a control group. An exposed group refers to those users who have viewed the advertisement (e.g., branded

content). A control group refers to those users that have not viewed the advertisement. A control group may also be referred to a non-exposed group or a non-exposed control group.

Conventional approaches for determining control groups (e.g., non-exposed group, non-exposed control group) in an advertisement campaign study (e.g., attitudinal shift studies, survey lift studies, survey lift analysis ,etc.) that measure attitudinal response to advertisements across the content item service are based upon hold-back mechanisms. Hold-back mechanisms provide a forced-view of an advertisement delivered via a reserve or auction system and is outside of any user selection or control. Correspondingly, the hold-back mechanism can cause users viewing the same content to not be shown the advertisement. Thus, the hold-back mechanisms allows for straight-forward generation of exposed groups and control groups (i.e., non-exposed groups) for the advertisement campaign study.

However, for "organic" views, where users choose to watch content not delivered via auction or advertisement systems, hold-backs are not possible. For instance, when a user selects a video to watch, the content item service cannot block that view in order to establish a hold back. Moreover, for "organic" views, all users who view the content including the advertisement are exposed to the advertisement. The advertisement cannot be held back from a view of the content. Thus, an ongoing problem with "organic" views and designing survey lift studies is identifying control groups for such studies. In other words, a gap exists for content uploaded to a content item service that encourages viewers to consider, review or buy products/services, but was not delivered in the form of an auction or reserve advertisement.

A mechanism for modeling statistically significant propensity controls is proposed. The proposed mechanism establishes a statistically significant control model for organic viewers in order to measure attitudinal shifts and lift from the viewing population. On a high level,

audience overlap is measured to find channels that are "similar" to the channels in an organic video campaign. A channel may refer to data content available from a common source or data content having a common topic or theme. The data content can be digital content chosen by a user, digital content made available by a user, digital content uploaded by a user, digital content chosen by a content provider, digital content chosen by a broadcaster, etc. For example, a channel X can include videos Y and Z. A channel can be associated with an owner, who is a user that can perform actions on the channel.

Different activities can be associated with the channel based on the owner's actions, such as the owner making digital content available on the channel, the owner selecting (e.g., liking) digital content associated with another channel, the owner commenting on digital content associated with another channel, etc. The activities associated with the channel can be collected into an activity feed for the channel. Users, other than the owner of the channel, can subscribe to one or more channels in which they are interested. Once a user subscribes to a channel, the user can be presented with information from the channel's activity feed. If a user subscribes to multiple channels, the activity feed for each channel to which the user is subscribed can be combined into a syndicated activity feed. Information from the syndicated activity feed can be presented to the user.

Referring back to the mechanism for establishing a statistically significant control model for organic views in order to measure attitudinal shifts and lift from the viewing population, the mechanism can use "active" subscribers/viewers of the identified "similar" channels as the non-exposed control group, after filtering out those users who watched the organic videos. The resulting channels that share viewers can be filtered by channel topicality (e.g., electronics reviews, beauty tips, etc.) and channel size (subscribers within a standard deviation). Viewers

are filtered by demographic, technographic and psychographic traits to align with that of the exposed groups.

Figure 1 depicts a flow diagram of a method for modeling statistically significant propensity controls. First, at step 101, channels of a content item service are identified, where the identified channels are associated with an advertisement campaign that utilizes organic videos. For example, there may be N channels (C1~Cn) in an advertisement campaign that contribute V1, V2, … Vn views, respectively, to the advertisement campaign. In total, there are V views ($V = \sum_{i=1}^{n} Vi$) of the advertisement campaign. X users are needed for the control group (i.e., non-exposed group) to conduct an attitudinal shift study (e.g., survey lift analysis). X can be determined by the total number of views and/or the expected number of responses.

At step 102, for each identified channel of the advertisement campaign, similar channels of the content item service are selected that are not in the advertisement campaign. Z channels may be chosen as the "most similar" channels for each advertisement campaign channel Z may be a number that is a balance between picking too few channels (picking from one channel may be biased) and too many channels. A data store of the content item service may be maintained that lists each channel of the content item service as well as those channels of the content item service that have an audience overlap with the particular channel. Audience overlap may refer to users having overlapping views of content items corresponding to both channels.

Then, at step 103, channels from the identified similar channels may be selected. The similar channels may by selected by applying one or more filters based on topicality and size. For example, the selected similar channels may have similar topicality to the identified advertisement campaign channel. Topicality may refer to a topic of the channel. Topics may include, but are not limited to, beauty, reviews, tips, and so on.

In addition, the selected similar channels may satisfy additional size criteria. The additional size criteria may include having enough active subscribers in the last 30 days (or any other determined number of days) and/or having enough active subscribers that are also active subscribers to the advertisement campaign channel. An active subscriber may refer to a user that has subscribed to the channel within a determined number of days (e.g., last 30 days) and/or any subscribed user to the channel that has viewed a content item from the channel within a predetermined number of days (e.g., last 30 days).

In one example, if three similar channels (e.g., Z =3) are being chosen for an advertisement campaign channel, the three channels may need to have enough subscribers in the last 30 days so that the number is subscribers is greater than $0.33 * (Vm/V)*X$ active subscribers. Furthermore, within the list of similar channels for the advertisement campaign channel, the three selected similar channels should have the highest (subscriber_pairs/total_active_subscribers) number.

Then, at step 104, for each identified channel of the advertisement campaign, subscribers of the selected similar channels are filtered that have viewed the organic videos of the advertisement campaign. This may be accomplished by referencing watch lists corresponding to sets of users maintained by the content item service. Subsequently, at step 105, those subscribers of the selected similar channels that were not filtered out at step 104 are then identified as candidates for a non-exposed control group candidate pool for the advertisement campaign study.

At step 106, for each selected similar channel, a determined number of the subscribers of the selected similar channels that were identified as candidates at step 104 are randomly added to a non-exposed control group candidate pool for the advertisement campaign. The determined

number may be based on a proportion of the number of views contributed by the identified similar channel to the total views provided by all of the selected similar channels. The determined number from each selected similar channel may further be based on the number of selected similar channels and the total number of non-exposed control group users needed for the advertisement campaign study.

For example, for each identified channel C_m of the advertisement campaign, $(V_m/V)*X$ non-exposed users may be needed from its similar channels. As a result, any channels that delivered more views contribute more non-exposed users to the non-exposed user control group, which makes the non-exposed group similar to the exposed group.

In a more detailed further example, the following provides an overview of how many control users may be needed from each similar channel when three similar channels (i.e., $Z = 3$) are selected for each identified advertisement campaign channel having channels C_1, C_2, …, C_n.

For channel C_1, having views V_1, the number of control users required for that channel C_1 may be $(V_1/V) * X$. Moreover, for channel C_1, the number of control users from each of similar channels 1, 2, and 3 may be $0.33 * (V_1/V) * X$ for each similar channel 1, 2, and 3.

For channel C_2, having views V_2, the number of control users required for that channel C_2 may be $(V_2/V) * X$. Moreover, for channel C_2, the number of control users from each of similar channels 1, 2, and 3 may be $0.33 * (V_2/V) * X$ for each similar channel 1, 2, and 3.

For channel C_3, having views V_3, the number of control users required for that channel C_3 may be $(V_3/V) * X$. Moreover, for channel C_3, the number of control users from each of similar channels 1, 2, and 3 may be $0.33 * (V_3/V) * X$ for each similar channel 1, 2, and 3.

At step 107, a non-exposed control group is generated for the advertisement campaign study. The non-exposed control group may be generated using the non-exposed control group

candidate pool created at step 106. In some cases, users may be selected from the non-exposed control group candidate pool based on demographic, technographic, and/or psychographic data in order to align with similar data from the exposed control group of the advertisement campaign study.

In situations in which the systems discussed herein collect personal information about users, or may make use of personal information, the users may be provided with an opportunity to control whether the content item service can collect user information *(e.g.,* information about a user's social network, social actions or activities, profession, a user's preferences, or a user's current location), or to control whether and/or how to receive content from the content server that may be more relevant to the user. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over how information is collected about the user and used by the content item service.

The above methodology detailed above for Figure 1 results in a reliable method for determining control groups based upon viewer and channel propensity in views not originated from hold-backs. In a test of trials using this methodology, control neutrality was determined to be statistically significant at the .95 confidence level.

The mechanism described herein allows for modeling statistically significant propensity controls in survey lift studies. Because the mechanism allows for creation of statistically significant non-exposed control groups for advertisement campaign studies, the impact of brand

recall after exposure to branded content can be accurately measured by a content item service. This accurate measurement of brand recall and/or attitudinal shift helps advertisers understand campaign performance. Furthermore, the methodology detailed herein results in control neutrality of the non-exposed control group, resulting in statistically-significant data generated by the advertisement campaign study. This statistically-significant data drives better quality control and product quality at the content item service.

ABSTRACT

A mechanism for modeling statistically significant propensity controls in survey lift studies is disclosed. The proposed mechanism establishes a statistically significant control model for organic viewers in order to measure attitudinal shifts and lift from the viewing population. On a high level, audience overlap is measured to find channels that are similar to the channels in an organic video campaign. Active subscribers of the identified similar channels are then used as the non-exposed group, after filtering out those who watched the organic videos. The resulting channels that share viewer audiences are filtered by channel topicality (e.g., electronics reviews, beauty tips, etc.) and channel size (subscribers within a standard deviation). Viewers are filtered by demographic, technographic and psychographic traits to align with that of the exposed groups.

Keywords: video, content, advertisement, survey, lift analysis, propensity control model, attitudinal shift, control group, non-exposed group, exposed group, organic view, organic video, overlapping viewers
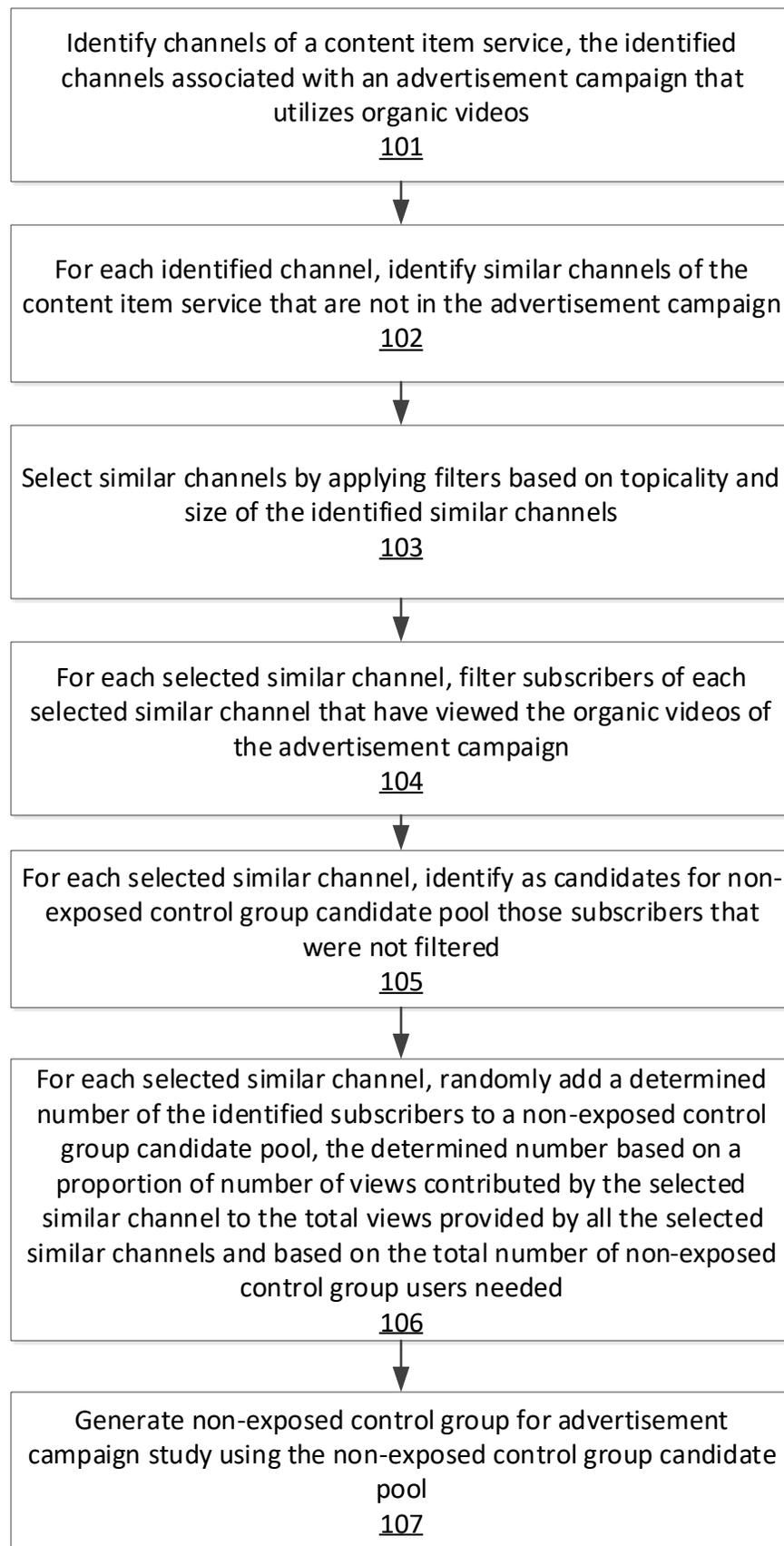
Identify channels of a content item service, the identified channels associated with an advertisement campaign that utilizes organic videos
101

For each identified channel, identify similar channels of the content item service that are not in the advertisement campaign
102

Select similar channels by applying filters based on topicality and size of the identified similar channels
103

For each selected similar channel, filter subscribers of each selected similar channel that have viewed the organic videos of the advertisement campaign
104

For each selected similar channel, identify as candidates for non-exposed control group candidate pool those subscribers that were not filtered
105

For each selected similar channel, randomly add a determined number of the identified subscribers to a non-exposed control group candidate pool, the determined number based on a proportion of number of views contributed by the selected similar channel to the total views provided by all the selected similar channels and based on the total number of non-exposed control group users needed
106

Generate non-exposed control group for advertisement campaign study using the non-exposed control group candidate pool
107

FIG. 1