Evaluating TV Ad Campaigns Using Set-Top Box Data

Sundar Dorai-Raj, Yannet Interian, and Dan Zigmond Google, Inc.

Abstract

Google has developed new metrics based on set-top box data for predicting the future audience retention of TV ads. This paper examines how to use these metrics to judge the effectiveness of TV ad campaigns. More specifically, we analyze how these metrics can inform future campaign targeting and placement goals.

Introduction

In recent years, there has been an explosion of interest in collecting and analyzing television set-top box (STB) data (also called "return path" data). As US television moves from analog to digital signals, digital set-top boxes are increasingly common in American homes. Where these STBs are attached to some sort of return path, this data can be aggregated and licensed to companies wishing to measure television viewership. For example, Google aggregates data, collected and anonymized by DISH Network L.L.C., describing the precise second-by-second tuning behavior from television set-top boxes in millions of US households. This data can be combined with detailed airing logs for thousands of daily TV ads to estimate second-by-second fluctuations in audience during TV commercials (Zigmond and Lanning, 2008).

These data hold the promise of providing accurate measurement for much of the niche TV content that eludes current panel-based methods. But in addition to using these data for raw audience measurement, it is possible to make more qualitative judgments about the content – and specifically the advertising – on television. Google has developed a measure of audience retention based on STB data that can be used to predict future audience response for TV ads (Zigmond, 2009a and Zigmond et al, 2009b).

This paper will look at how this new retention metric can be applied to measure the effectiveness of TV ad campaigns.

Retention Scores

Raw measures of audience tuning behavior during TV ads can be useful in evaluating TV ads. However, we have found that these metrics are highly influenced by extraneous factors such as the time-of-day, day-of-week, and the network on which the ads were aired. These are nuisance variables and make direct comparison of such measures very difficult.

Rather than using these measures directly, we have developed a model for normalizing the scores relative to expected tuning behavior (Zigmond et al, 2009b). We do this by using a statistical model to estimate the "expected" tuning behavior during a given ad spot (based on known influencing factors like time-of-day, day-of-week, etc.), and subtract from this the observed tuning behavior during a specific ad airing. We then score ads or campaigns by looking at the percentage of airings in which this residual (ie, the expected minus the actual) exceeds the median. We call this quantity the "retention

score" because it attempts to capture the audience retention directly attributable to an ad or campaign itself.

Retention and Ad Campaigns

We have started using retention scores for a variety of applications at Google. These scores are made available to advertisers, who can use them to evaluate how well their campaigns are retaining audience. This can be a useful proxy for the relevance of their ads in specific settings.

For example, Figure 1 shows the retention scores for an automotive advertiser, compared with the average scores for other automotive companies advertising on television with Google. Separate scores were calculated for each network on which this advertiser aired. We can see not only significant differences in the retention scores for these ads, but also differences in the relative scores compared against the industry average. On the National Geographic Channel, for example, this advertiser's retention scores exceed those of the industry average by a significant margin. This sort of analysis can be used to suggest ad placements where viewers seem to be more receptive to a given ad.



Figure 1. This plot shows retention scores for ads run by an auto manufacturer (black bars) to their competitors' ads (gray bars). Some networks have better scores than others, which provide important feedback to the advertisers. The length of the bar represents a 90% confidence interval on the score.

Figure 2 shows another application, this time using retention scores calculated within specific demographic clusters to evaluate how receptive viewers in that cluster are to a four different ads for an online service. Here, viewers with an interest in technology (darker bars) often have higher retention than those without that interest, particularly for the longer form of creative A. This suggests that this version of the ad may be especially appealing to those viewers.

This paper will include several more specific case studies like the ones above, in which retention scores were able to help an advertiser evaluate live TV campaigns and make specific decisions about future TV spending.



Figure 2. This plot shows retention scores for two different ads, each of two different lengths, for an online service. The colored bars show the retention scores for each of four demographic categories. The length of the bar represents a 90% confidence interval on the score.

Conclusions

Retention scores are an important new addition to existing advertising metrics. This paper will show how advertisers can use these scores to evaluate running campaigns and to guide future spending. In the long run, we hope this new style of metric will inspire and encourage better and more relevant advertising on television.

Bibliography

- Zigmond, Dan, "Do Viewers Care? Understanding the impact of creative on TV viewing behavior" at <u>Re:think 2009: The ARF Annual Convention</u>, April 1, 2009.
- Zigmond, Dan and Steve Lanning, "Learning from Tuning: Developing New Ad Metrics from Set-Top Box Data" at <u>Audience Measurement 3.0</u>, June 25, 2008.
- Zigmond, Dan, Sundar Dorai-Raj, Yannet Interian, Igor Naverniouk, "Measuring Advertising Quality on Television: Deriving Meaningful Metrics from Audience Retention Data" Journal of Advertising Research, vol. 49 (2009), pp. 419-428.