How Surfers Watch: Measuring audience response to video advertising online

Sundar Dorai-Raj
Google, Inc
1600 Amphitheatre Pkwy
Mountain View CA 93043
sdorairaj@google.com

Dan Zigmond
Google, Inc
1600 Amphitheatre Pkwy
Mountain View CA 94043
djz@google.com

ABSTRACT

For several years, Google has been analyzing television set-top box data to measure audience response to specific TV ads. This paper presents how similar techniques can be applied to online video advertising on YouTube. As more and more video programming is made available online, it will become increasingly important to understand how to engage with online viewers through video advertising. Furthermore, we find that viewing behavior is even more effected by specific video ad creatives online than it is on TV. This suggests that online viewing can become a valuable source data on viewer response to video ad creatives more generally.

1. INTRODUCTION

Google analyzes anonymized television set-top box data for thousands of TV ads aired each day [5] and uses this data to measure how these ads affect viewing behavior [2, 4]. Although many factors lead viewers do tune away from TV ads, we find that a significant portion of the variance in tune-away rates can be attributed to the ad creative itself. Using statistical models to isolate this creative effect and to score ads based on how likely they seem to be to retain their initial viewing audience.

In the summer of 2009, we began applying similar techniques to viewing logs from YouTube, the online video sharing site founded in 2005 and acquired by Google in 2006. Users worldwide watch over a billion videos a day on YouTube, and upload hundreds of thousands of videos daily. YouTube incorporates several ad formats and generates over a billion ad impressions each week. Since 2008, some YouTube videos have included in-stream ads, video ads interspersed into the online videos. Because of their similarity to TV ads, these in-stream ads are the focus on the work described here.

In addition, we introduced an experiment on YouTube at the end of 2009 to test skippable in-stream ads. The experiment allowed users to skip in-stream ads at 0, 5, 10, and 15 seconds, and also included a control, which did not allow skipping. The goal of this experiment was to determine if allowing a user to skip an ad would decrease video abandonment.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ADKDD'10, July 25, 2010, Washington D.C., USA.
Copyright 2010 ACM 978-1-4503-0221-0 ...$10.00.

In this paper, we discuss how users interact with in-stream video ads on YouTube. Section 2 introduces the metric by which we measure ad abandonment. Section 3 describes methods we use to measure the effects due to partner (or network) and creative. Section 4 compares the effects across different data sources.

2. MEASURING AUDIENCE RESPONSE ONLINE

Although online viewers do not "tune away" in the same way that TV viewers do, they do sometimes abandon ads by closing the viewing window or otherwise ending their online session. This online ad abandonment can be modeled using the same statistical techniques we have previously applied to TV tune-away rates.

In [1], we defined IAR, or initial audience retained, as

$$\text{IAR} = \frac{\text{Audience that viewed the whole ad}}{\text{Audience at the beginning}}. \quad (1)$$

Using IAR as the response in a logistic regression model, we are able to account for certain known features, such as ad length, that effect the probability a viewer will tune away. However, two features we would like to understand better are the effect due to the creator of the underlying video programming (the programmer on TV or partner online) and the creative effect. While the partner effect shows that some channels are better for showing ads than others, the creative effect shows that some ads are consistently better than others.

3. MEASURING THE PARTNER AND CREATIVE EFFECTS

To measure both the partner and creative effects, we fit a series of sequential logistic regression models of the form:

$$\begin{aligned} \log\left(\frac{IAR}{1-IAR}\right) &\sim \text{(Intercept)} && (2)\\ &\sim M_0 && (3)\\ &\sim M_0 + \text{partner\_id} && (4)\\ &\sim M_0 + \text{partner\_id} + \text{creative\_id}, && (5) \end{aligned}$$

where each model in (2)-(5) differs from the previous model by only one term, and M_0 is a set of known features defined in Table 1. The "partner_id" contains a unique id for every partner and the "creative_id" contains a unique if for every creative. The null model (or intercept-only) model given by (2) provides a baseline model from which we will make all subsquent comparisons.

YouTube In-Stream	weekday pre-/mid-/postroll country ad length video length ad length × video length
YouTube Skippable In-Stream	ad length video length skip time skip time × ad length
TV	daypart weekend ad duration

Table 1: The sets of known features used to describe YouTube in-stream ads, YouTube skippable in-stream ads, and TV ads. Coefficients for each feature are fit using logistic regression with IAR as a response.

From each model in (2)-(5) we calculate the model deviance given by

$$D(\mathbf{y}, \mathbf{p}) \propto -2 \sum_{i=1}^{n} y_i \log \frac{p_i}{1-p_i} + \log(1-p_i), \qquad (6)$$

where $\mathbf{y}$ is a length n vector of observed IAR and $\mathbf{p}$ is a vector of the expected IAR from one of the models in (2)-(5) [3]. For logistic regression, (6) may be thought of as the variance of $\mathbf{y}$ described by $\mathbf{p}$. This implies that relative changes in deviance from one model to the next may be interpreted as the additional percentage of variance explained by adding a new feature.

The null deviance D_0, which is determined from (2), represents the total variance in the data. Then to measure the partner and creative effect, we determine the deviances D_M, D_p, and D_c using models (3), (4), and (5), respectively. The percentages of variance explained by the partner and creative are given by

$$\begin{aligned} \text{\% deviance explained by partner} &= \frac{D_p - D_M}{D_0} \times 100\% \\ \text{\% deviance explained by creative} &= \frac{D_c - D_p}{D_0} \times 100\%. \end{aligned}$$

4. COMPARING CREATIVE EFFECTS

We found that the creative's impact on audience retention appears higher on YouTube than on TV, with creatives explaining about 24% of the variance on YouTube compared with just 8% on TV, as shown in Figure 1 below. Moreover, when users were allowed to skip ads, we found that the creative effect decreased to 10%. This result is not surprising since users who skip ads do not usually stick around long enough to figure out whether the ad is appealing or not.

Conversely, the creator of the underlying video programming (the programmer on TV or partner online) impact appears smaller online than on TV. This suggests that online viewers may be less creatures of habit than TV viewers.

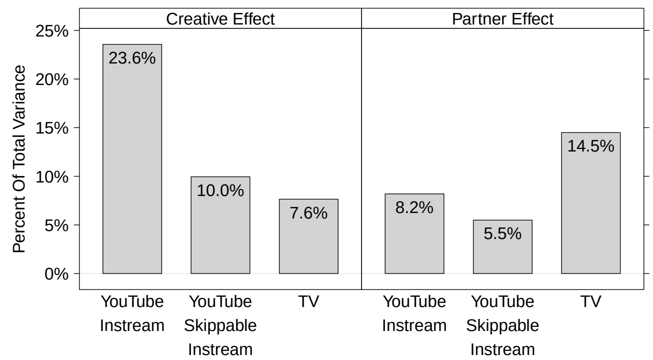

Figure 1: The percentage of variance explained by the creative effect (left) and partner effect (right) for YouTube in-stream ads, YouTube skippable in-stream ads, and TV ads. The creative effect is greatly diminished when ads become skippable on YouTube implying that creative differences among ads are less discernible if you a user is allowed to skip it.

YouTube allows a variety of in-stream ad placements, including pre-roll ads (played before the video begins), mid-roll ads (interspersed during the video), and post-roll ads (added at the end). For each ad, we calculated the audience retention as a percentage of the initial audience retained (IAR), and Figure 2 shows the distribution of IAR value of each of these three categories of ad placement. Unsurprisingly, we found that viewers were much more likely to watch pre-roll ads than mid-roll or post-roll, as shown in Figure 2.

Across all three placements, however, YouTube viewers were much more active in their viewing than most TV viewers. Where audience retention remained above 90% for virtually all TV ads considered, YouTube ads showed much greater variety, with more than 40 percentage points difference between the best- and worst-performing ad creatives. This further confirms the results shown in Figure 1 that viewers are more responsive to specific ad creatives in their online viewing behavior than they are when watching TV.

Because many online viewers seem to prefer shorter videos, we looked specifically at ad abandonment rates for YouTube pre-roll ads on videos less than 22 minutes. The distribution of IAR values for both 15- and 30-second ads for these shorter videos is shown in Figure 3, with shorter videos in light gray and longer videos in black.

We found no significant difference in overall audience retention rates for 15-second ads to 30-second ads on these shorter videos, suggesting that viewers were equally accepting of both. However, we found that for the shorter (15-second) ads, abandonment rates were relatively constant regardless of the length of the video, while for 30-second ads there was a clear decrease in abandonment as the video length increased. In other words, it seems users are willing to tolerate longer pre-roll ads for longer videos, but not for shorter video, but that 15-second ads are well tolerated regardless.

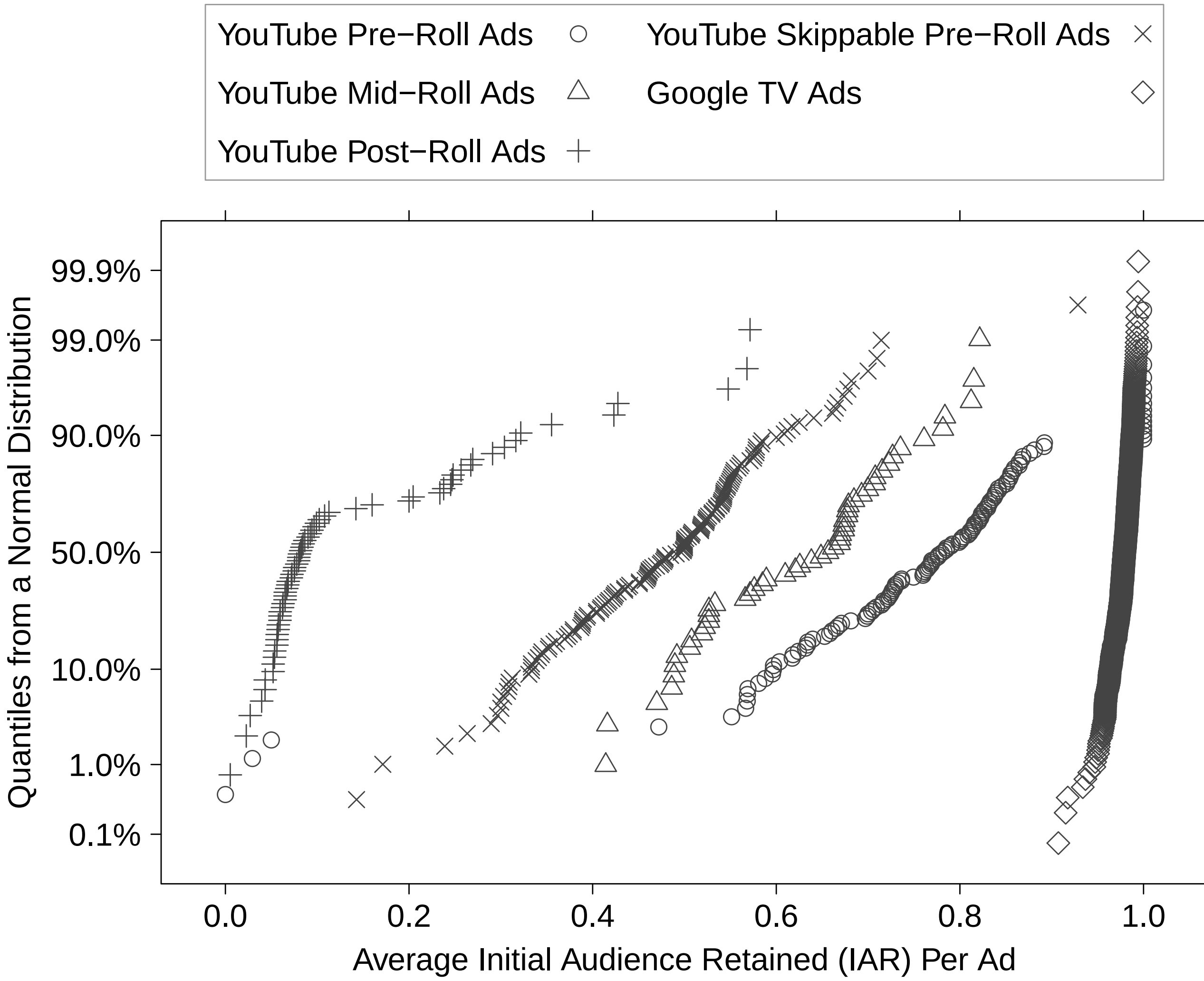

Figure 2: The IAR for pre-roll, mid-roll, and post-roll in-stream ads along with skippable ads and TV ads. TV ads (⋄) have a fairly high IAR since most TV viewers are very passive. Post-roll in-stream ads (+) have the worst IAR since most viewers would rather not sit through an ad after the video has finished. Skippable ads (×), however, have an IAR worse than a mid-roll in-stream ads.

5. CONCLUSIONS AND FUTURE WORK

The work presented in this paper is preliminary and considerably more can be done to shown causality between an ad's creative effect and abandonment. We currently have better models in place to help predict tune away both on TV and YouTube using more sophisticated machine learning algorithms. We hope to update our work in this area in the future.

With so much video programming now available online, the advertising industry needs ways of measuring the response rate of users to online video advertising. Advertisers are still exploring what sorts of advertising works online and whether lessons learned from video ads on TV can be applied to this new medium. Traditional online metrics like click-through rates offer one approach, but in many cases this will not be applicable to video ads that are not intended to solicit an immediate response. Audience retention rates, which Google has applied to TV advertising for several years, offer an alternative approach.

6. REFERENCES

[1] Y. Interian, S. Dorai-Raj, I. Naverniouk, P. Opalinski, Kaustuv, and D. Zigmond. Ad quality on tv: Predicting television audience retention. In *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*, July 2009.

[2] Y. Interian, Kaustuv, I. Naverniouk, P. Opalinski, S. Dorai-Raj, and D. Zigmond. Do viewers care? understanding the impact of ad creatives on tv viewing behavior. In *Re:Think09*, April 2009.

[3] P. McCullagh and J. Nelder. *Generalized Linear Models*. Chapman and Hall, London, 1989.

[4] D. Zigmond, S. Dorai-Raj, Y. Interian, and I. Naverniouk. Measuring advertising quality on

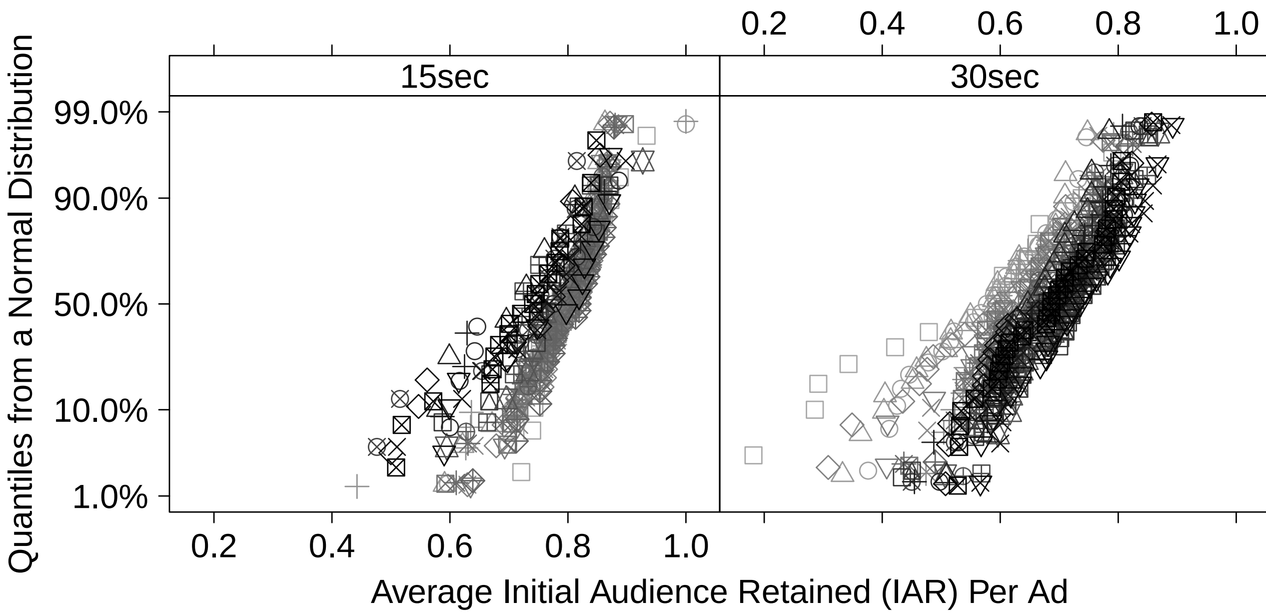

Figure 3: The IAR of ads that appear on videos under 22 minutes long. The lighter shades represent shorter videos, while the darker shades represent longer videos. For 30-second ads, we can see a fairly steady decline in IAR as the videos get shorter suggesting users are more tolerant of in-stream ads on longer videos.

television: Deriving meaningful metrics from audience retention data. *Journal of Advertising Research*, 49:419–428, December 2009.

[5] D. Zigmond and S. Lanning. Learning from tuning: Developing new ad metrics from set-top box data. In *Audience Measurement 3.0*, June 2008.