The influence of visual salience on video consumption behavior: A survival analysis approach

Rafael Huber*
University of Basel
Missionsstrasse 60
4055 Basel, Switzerland
+41 61 267 35 31
huber.rafael@gmail.com

Benjamin Scheibehenne
University of Basel
4055 Basel, Switzerland
+41 61 267 35 31
benjamin.scheibehenne@unibas.ch

Alexandre Chapiro
ETH Zürich
Universitätstrasse 6
8003, Zürich, Switzerland
alexandre.chapiro@disneyresearch.com

Seth Frey
Disney Research Zürich
Stampfenbachstrasse 48
8006 Zürich, Switzerland
+41 44 632 7360
seth@disneyresearch.com

Robert W. Sumner
Disney Research Zürich
Stampfenbachstrasse 48
8006 Zürich, Switzerland
+41 44 632 7360
sumner@disneyresearch.com

ABSTRACT

In an increasingly competitive media environment, producers of online content need analytics that can predict the success of a video. In recent years the field of visual computation has produced a variety of mathematical models that quantify an image's salience, that is, its potential to capture attention. To test how a video's content might predict its success, we applied the standard saliency model of Itti, Koch, and Niebur [1] to more than 1000 video clips that were broadcast on a large video streaming website. We also obtained fine-grained data on the viewership of these clips. Based on a survival analysis, we find that people prefer more salient videos. The results were robust towards the inclusion of other predictors such as the genre of the video, but not to video length, which remains correlated with salience even after comparing videos only within show and genre. Our analyses suggest that visual salience provides an objective and easy-to-compute supplement to previously suggested predictors of video consumption behavior.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems – *human factors, human information processing*. I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding – *intensity, color, photometry, and thresholding*.

General Terms

Human Factors.

WebSci '15, June 28 - July 01, 2015, Oxford, United Kingdom
ACM 978-1-4503-3672-7/15/06.
http://dx.doi.org/10.1145/2786451.2786507

Keywords

salience, visual salience, survival analysis, predictive analytics, video consumption behavior

1. INTRODUCTION

It is important to understand and predict what drives the success of video clips. To improve prediction accuracy, one promising approach is to examine objective properties of video content for hints about its success in front of an audience. Starting with basic psychological theories of visual perception, we test whether video consumption statistics can be predicted with quantitative measures of visual salience, as defined in terms of the levels and variability in the brightness, orientation, and color of video frames.

2. METHOD

Our analysis is based on usage data from a website that allowed users to browse clips from popular television shows. On average, videos were short clips of about 3 minutes. Here, we focus on users' probabilities of navigating away from a video before its conclusion, and how that probability is affected by the salience of the raw video that they watched.

2.1 Estimating Salience

Itti, Koch and Niebur [1] proposed a quantitative, computationally tractable, and empirically-validated model of salience that takes objective visual properties of the environment such as color, intensity, and orientation features as input [2]. This so-called IKN estimator is used widely, and it regularly features as a benchmark for evaluating alternative experimental measures [3]. Salience was calculated by means of the Matlab Saliency Toolbox, a ready-to-use package with implementations of the "IKN" measure. In its original form, the IKN salience assigns to each pixel a vector representing color, brightness, and orientation dimensions of visual salience. Since our aim was to obtain a single comprehensive salience value for each video, we summed the computed per-pixel saliency values across all pixels within each frame. We then computed two alternative salience measures, one based on the mean of the IKN (across all frames within each

* The authors wish to thank Steven Poulakos, Aljoscha Smolic, and Maurizio Nitti for their assistance and helpful comments. Author RH conducted this work with the support of Disney Research. For an extended version of this manuscript, contact author SF.

video) and the other based on its standard deviation. Both measures have a straightforward interpretation: The mean expresses the general intensity or “attention grabbing” potential that presumably demands high resources in tonic, sustained attention and the standard deviation is a proxy for the variability of the salience across a video that presumably demands high resources in phasic attention and/or alertness.

The median of the mean salience across all videos was 31.62 with a range from 19.26 to 43.19. The median of the standard deviation-based salience was 8.41 with a range from 4.11 to 14.99. Calculating both salience measures for each video was computationally intensive as it involved processing 1,007 videos for which raw data was available frame-by-frame, and that were watched at least once. Videos were encoded at 24–30 fps with the MPEG-4 Part 14 (MP4) codec. On an Intel® Core™ i7-3930K CPU @ 3.20GHz, computation took approximately 150 ms per frame of video, or approximately two weeks for the entire dataset.

2.3 Statistical Analysis

To test the relationship between video salience and video consumption behavior, we estimated each video's hazard rate, defined as the probability of navigating away from it before it has ended, as a function of proportion viewed, for high- and low-salience videos, separately for two measures of salience. The statistical method of survival analysis provides a feasible tool for this analysis [4]. Here, “survival” means that users have not yet clicked away from the video. Thus, the longer one watches a video, the longer this person “survives.”

When conducting a survival analysis, one tries to approximate the survival function $S(t)$, that is the probability of an individual's "survival" T exceeding time t: $S(t) = P(T > t)$. Theoretically, this probability is best described by the true underlying survival function $S(t)$. Because the theoretical function is not known, one has to rely on an empirical approximation called the Kaplan-Meier (KM) method. Here, we defined “stop watching a video” as the event of interest and the percentage $p = [0,1]$ of a video watched until an event occurred as the time of survival. Because not all videos had the same duration, percentage watched can be regarded as a comparable metric across videos. We labeled videos either “high” or “low” salience based on whether their ratings were above or below the median across all videos. We made an overall statistical comparison of the two empirical KM survival curves using the log-rank test (“Mantel-Haenszel test”).

3. RESULTS

The core dataset of our analysis consisted of 81,041 data points, with behavior from 31,671 unique viewers. Most viewers watched just one video (75%) and 2% of viewers watched ten or more.

To test the relationship between video consumption behavior and mean- and standard deviation-based measures of salience, we fit KM curves separately for videos with “high” and “low” values of salience, and then compared them with log-rank tests. The comparison of the KM estimates for the survival curves (“high” versus “low,” see Figure 1) showed a significant difference for the mean-based measure of salience ($\chi^2 = 1{,}261$, $df = 1$, $p < 0.001$) and a marginally significant trend for the standard deviation-based measure ($\chi^2 = 3.7$, $df = 1$, $p = 0.0535$).

To test the robustness of these results, we also conducted additional analyses that included additional contextual predictors above and beyond salience. These are reported in detail in the full manuscript.

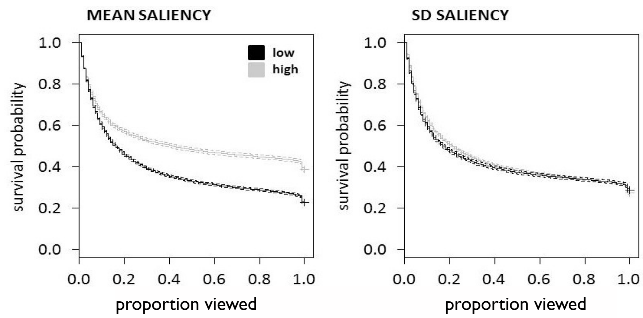

Figure 1. Kaplan-Meier estimates for the mean and standard deviation (sd) of salience, separately for high (grey) and low (black) values of the measures as defined by a median split. Videos with high and low saliency show differences in the rate at which viewers drop out of videos, particularly the mean saliency over the course of a video. X-axis defines dropout rates in the sense that 100% of viewers completed 0% of all videos (far left of curves) and approximately 20%–40% of viewers completed 100% of all videos (far right).

4. DISCUSSION

We defined the success of a video as the amount of it that was completed (in both percentage and seconds) before a viewer navigated away, and we suggest an objective and easy to compute measure – visual salience – as a promising predictor of movie success. We find that our simple measure of low-level psychological salience can predict how likely viewers are to persist through online videos in a browsing environment.

Our results suggest that low-level psychophysical variables may play an important role in sustaining attention for the duration of the short video clips that we analyzed here. These results dovetail with past research showing that attention processes play an important role in consumer behavior [5, 6]. IKN saliency is not susceptible to observer effects or to the vagaries of self-reports, it is simple to implement, and it can be readily applied to online video content, where conventional, high-level predictors for motion pictures success may be less relevant. Our work presents salience as a promising variable for predictive analytics and marketing applications.

5. REFERENCES

[1] Itti, L. et al. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions pattern analysis and machine intelligence*. 20, 11 (1998), 1254–1259.

[2] Walther, D. and Koch, C. 2006. Modeling attention to salient proto-objects. *Neural networks*. 19, 9 (2006), 1395–1407.

[3] Borji, A. and Itti, L. 2013. State-of-the-art in visual attention modeling. *IEEE Transactions pattern analysis and machine intelligence*. 35, 1 (2013), 185–207.

[4] Kleinbaum, D.G. 1998. Survival Analysis, a Self-Learning Text. *Biometrical Journal*. 40, 1 (1998), 107–108.

[5] Pieters, R. and Warlop, L. 1999. Visual attention during brand choice: The impact of time pressure and task motivation. *Intern. J. of Research in Marketing*. 16, (1999), 1–16.

[6] Pieters, R. et al. 1999. Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory. *Journal of Marketing Research* 36, 424–438 (1999).