This Image Intentionally Left Blank: Mundane Images Increase Citizen Science Participation

Alex Bowyer
alex.bowyer@zooniverse.org
University of Oxford / Zooniverse

Veronica Maidel
veronica@zooniverse.org
Adler Planetarium / Zooniverse

Chris Lintott
chris@zooniverse.org
University of Oxford / Zooniverse

Ali Swanson
ali@zooniverse.org
University of Oxford / Zooniverse

Grant Miller
grant@zooniverse.org
University of Oxford / Zooniverse

Abstract

Snapshot Serengeti is a Zooniverse crowdsourcing project that invites volunteers to identify animals present in automatically obtained photographs from the Serengeti. Our experiment explored the relationship between "blank" images (those containing no animals) and session length, theorising that, in line with "intermittent reward" theory, a larger blanks-to-animals ratio could increase engagement. Cohorts drawing from image pools with 0%, 20%, 40%, 60% and 80% blank images showed a decrease in mean session length as percentage blanks decreased; the effect size grows from small to moderate. The null hypothesis that all cohorts have an equal mean was rejected, confirming that the percentage of available blank images impacts upon session length. Our finding that removing blanks can decrease contribution will help other crowdsourcing projects increase engagement.

Introduction

The Zooniverse[1] is home to the Internet's largest, most popular and most successful citizen science projects. Its projects, each created for a specific research domain, consist of a web application that allow volunteers to perform classification tasks on subjects - typically those where humans perform better than computers, such as pattern recognition or transcription. The Snapshot Serengeti[2] project asks volunteers to examine a subject comprising 1-3 photographs and to identify the species, number and behaviour of any animals therein. The aggregate opinions of the volunteers are then used to determine the content of the subject (Hines 2015), which is fed back to the science team to enable studies in animal population & migrations (Swanson 2015). In order to maximise the speed with which classifications are collected, we would like to increase each user's contribution, measured in terms of session length - the number of subjects seen by a user in one sitting. Session length is calculated by looking at the time elapsed between classifications, and starting a new session whenever it reaches 30 minutes. Our research aims to discover if engagement can be increased by optimising the ratio of "blank" images (those containing no animals) to those with animals present. A recent volunteer survey ([Lintott] in prep) suggests that finding animals only occasionally is motivating. Research from the fields of psychology, gambling and computer games supports the idea that intermittent reward (Kendall 1974) can increase the contribution of an individual, especially when the ratio of rewards is variable (Ferster 1957, Yukl 1972, Delabbro 1998, Hopson 2001). In this context, a non-blank image occurring after a series of blank images could be considered a reward or "immediate payoff" (Kaufmann 2011).

Experiment Design & Implementation

Snapshot Serengeti normally presents a volunteer with a randomly selected subject from a large pool (about 70% of which are typically blank). The approach of our experiments is to intervene in this process and have certain cohorts of users select at random from controlled sets of subjects with certain characteristics. To manage the experiment and collect observations, we built a framework consisting of an analytics collector, Geordi[3], and a PlanOut-based experiment server[4]. This experiment consisted of 6 cohorts - a control group, and 5 experimental groups drawing from pools of previously analysed subjects containing 0%, 20%, 40%, 60% and 80% blanks. To preserve the fundamental user experience, we did not include a 100% cohort. Each user had a unique pool of 70 subjects (a size considerably larger than the mean session length of 23), and each user's participation in the experiment ended once their set was exhausted. The experiment ran for 3 weeks in July 2015 with 1,882 participants contributing 4,204 unique sessions.

Data Cleansing

In order to maintain a consistent user experience, "filler" non-experimental subjects were presented during server outages and during calculation of each user's subject pool. We wrote a script to discard all user sessions containing such filler subjects. Our script also ensured that only the first 1 - 70 experimental subjects in each session were

[1] http://www.zooniverse.org/
[2] http://www.snapshotserengeti.org/
[3] https://github.com/zooniverse/geordi/tree/master/geordi
[4] https://github.com/zooniverse/ZooniverseExperimentServer

counted. After cleansing, 1,506 user sessions were left for analysis.

Analysis & Findings

Initial analyses of the experimental data found that session lengths have an exponential distribution for each cohort, with most users having short sessions and only a few users having very long sessions. Although each cohort was centred on the intended percentage of blanks, within-group variance was large, because short sessions, presenting subjects in random order from the pool, have a very wide spread. This means that the intervention (exposing the users to a certain percentage of blanks) was not as accurate as expected. Nonetheless, mean session length does increase across experimental cohorts, as shown in Figure 1.

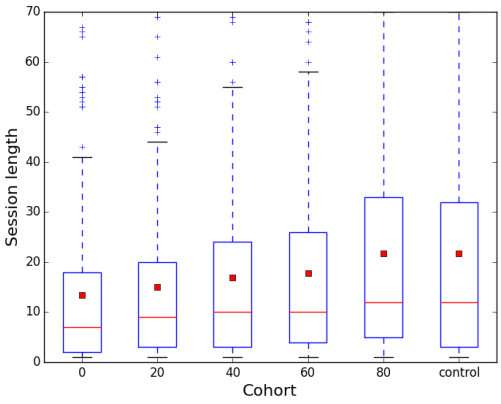

Fig. 1: Box-and-whisker plot showing sessions in each cohort

T-tests confirmed that the first four cohorts significantly differ from control; p was < 0.01 for 0%, 20% and 40%, and < 0.05 for the 60% cohort. We then calculated session length effect size for each experimental group relative to control, along with a confidence interval. Glass' Δ (Glass 1976) was used, as it does not assume equality of variances. As shown in Fig. 2, we found a decreasing negative effect size from moderate to small as % blanks increases.

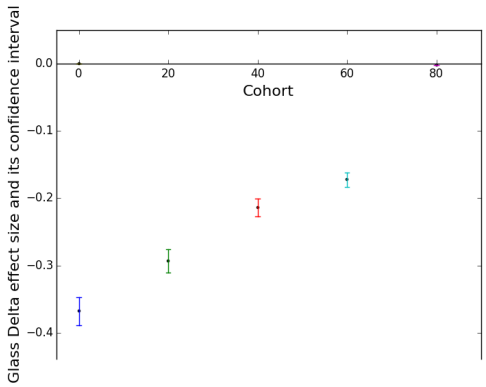

Fig. 2: Glass' Δ effect size of % blanks upon session length

The control cohort and the 80% cohort had similar session length, perhaps because control had 71% blanks. In order to validate the difference between cohorts, we proposed a null hypothesis: *"There are no differences in mean session length between the cohorts."* The Kruskal-Wallis test was used to test this, because it does not require a normal distribution or equal variances. The null hypothesis was successfully rejected, with $p < 0.05$. Dunn's post-hoc test was then performed to examine pairs of cohorts and see which ones differed from each other. The following pairs of cohorts were found to have significant differences: 0% vs. control, 20% vs. 80%, 0% vs. 80%. This confirms that session lengths are longer with ~80% blanks as opposed to 0% or 20% blanks.

Conclusion

The significant effect sizes and distinct means of the different cohorts show that reducing the percentage of blanks from 80% to 0% will shorten users' sessions. This supports the theory that a high percentage of blanks produces higher engagement because occasional animal sightings increase the feeling of reward. This research is of immediate value to other crowdsourcing projects: Removing "blank images" in an attempt to increase efficiency will likely reduce user participation; it is important to optimize task assignment for volunteer experience, not merely for efficiency. We plan to repeat the experiment with 5%-interval cohorts, to explore the 60%-90% range in more detail, and to discover if there is an optimal percentage of blanks that can have a positive effect size versus control, and result in a longer mean session length than control. In order to reduce the variance experienced within each cohort, we will explore controlling the order of subjects as the user progresses, keeping them closer to the target percentage. In future we will use the MICO platform (Aichroth 2015) to automatically identify blanks prior to human analysis.

Acknowledgements

Funding was provided by the EU project MICO (Media In Context) [grant no 610480, 2014-2017], SoCS (Social-Computational Systems) [award no: 1211094], the University of Oxford, and Adler Planetarium, Chicago.

References

Aichroth, Patrick, et al. "MICO-Media in Context." Multimedia & Expo Workshops (ICMEW), *2015 IEEE International Conference on*. IEEE, 2015.

Delabbro, P. & Winefield, A. "Poker-machine gambling: An analysis of within session characteristics." *British Journal of Psychology* 90.3 (1999): 469-490

Ferster, C. & Skinner, B. "Schedules of reinforcement." (1957): 13-14
Hines, G, et al. "Aggregating User Input in Ecology Citizen Science Projects." *Twenty-Seventh IAAI Conference*. 2015.

Glass, Gene V. "Primary, secondary, and meta-analysis of research." *Educational researcher* (1976): 3-8.

Hopson, J. "Behavioral game design." Gamasutra, April 27 (2001): 2001.
Kaufmann, N & Schulze, T. "Worker motivation in crowdsourcing and human computation." *Education* 17.2009 (2011): 6.

Kendall, S. "Preference for intermittent reinforcement." *Journal of the Experimental Analysis of Behavior* 21.3 (1974): 463.

Swanson, A, et al. "Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna." Scientific *data* 2 (2015).

Yukl, G et al. "Effectiveness of pay incentives under variable ratio and continuous reinforcement schedules." *Journal of Applied Psychology* 56.1 (1972): 19..