Using Screenshots to Predict Task Switching on Smartphones

Xiao Yang
Human Development and Family Studies
Pennsylvania State University
University Park, PA, USA
xfy5031@psu.edu

Nilam Ram
Human Development and Family Studies
Pennsylvania State University
University Park, PA, USA
nur5@psu.edu

Thomas Robinson
Departments of Pediatrics and Medicine
Stanford University
Stanford, CA, USA
tom.robinson@stanford.edu

Byron Reeves
Department of Communication
Stanford University
Stanford, CA, USA
reeves@stanford.edu

ABSTRACT

Mobile phone use is pervasive, yet little is known about task switching on digital platforms and applications. We propose an unobtrusive experience sampling method to observe how individuals use their smartphones by taking screenshots every 5 seconds when the device is on. The purpose of this paper is to incorporate the psychological process into feature extraction, and use these features to effectively predict the task switching behavior on smartphones. Features are extracted from the sequence of screenshots, gauging visual stimulation, cognitive load, velocity and accumulation, sentiment, and time-related factors. Labels of task switching behavior were manually tagged for 87,182 screenshots from 60 subjects. Using random forest, we demonstrate that we can correctly infer a user's task switching behavior from unstructured data in screenshots with up to 77% accuracy, demonstrating it is a viable option to use features of the screenshots to predict task switching behavior.

KEYWORDS

task switching; unobtrusive experience sampling; screenshots

CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.

ACM ISBN 978-1-4503-5971-9/19/05. https://doi.org/10.1145/3290607.3313089

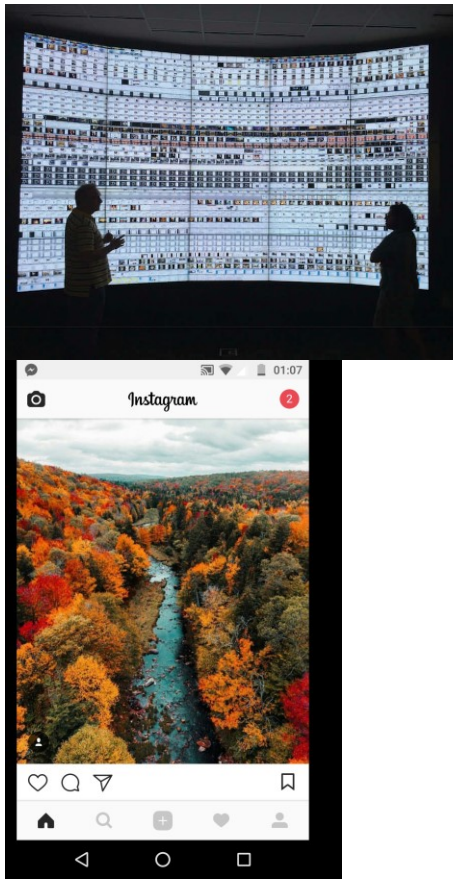

Figure 1: Data collection (top: two researchers were discussing this project while looking at 75 hours of screenshots from one subject, bottom: example of one screenshot).

INTRODUCTION

Life experiences are increasingly consolidated on one device – the smartphone. Individuals can now engage and accomplish a broad range of tasks on their smartphones, and it includes everything from communicating to gathering information to entertainment to shopping and much more [11]. An observational study of individuals' laptop use suggests that individuals switch from task to task very often, on average every 19 seconds [14]. Fine-grained study of task switching at this time-scale requires unobtrusive data collection given the impracticality of asking individuals to self-report on their task switching every few seconds. This paper illustrates new work wherein intensive, unobtrusive experience sampling is used to observe and predict individuals' *in-situ* task switching on smartphones. As shown in Fig. 1, continuous screenshots were collected without interfering with user activity on the smartphone.

Following prior research on how individuals switch among TV channels [13], switching among tasks and applications on smartphones is considered the behavioral output of a dynamic motivation processing system. Motivation can be categorized in two dimensions, an appetitive motivation that seeks novel media content and an aversive motivation that continues consuming the current content. Task switching has been associated with the activation of the appetitive motivation, and this is supported by both physiological measures and self-report, e.g., increased skin conductance levels were found consistently prior to switches [14], and risk takers (high on appetitive motivation and low on aversive motivation) were found to have the shortest length of content segment compared with other motivation types [15]. In parallel, limited capacity model of media processing [7] suggests that when the information load of current media exceeds cognitive processing capacity, the overload elicits aversive motivation to ease discomfort associated with cognitive load. Together these models suggest that task switching on smartphones can be predicted, at least in part, by information-based features of the media stimuli – what appears on the smartphone screen.

In this paper, we develop a prediction model for user's task switching on smartphones using a new data source and machine learning algorithms (random forests). Features of the media content that appears on individuals' smartphone screens are derived from sequences of screenshots obtained every 5 seconds that the device is in use and used to predict task-switches. Specifically, we explore the relative importance of measures of visual stimulation, cognitive load, velocity of stimulation and cognitive load, accumulation of visual stimulation and cognitive load, sentiment, and time-related factors for identifying when individuals will switch to new tasks or applications. Identification of important features supports future study of human-computer interaction, including design and delivery of media content that maximizes user engagement and individual differences in response to various kinds of media content. This paper illustrates a new data collection and analytical paradigm that will facilitate new knowledge about individuals' *in situ* smartphone use and how features of smartphone applications influence patterns of use.

Figure 2: Example of labeling task switching (screenshot with red border) from a sequence of screenshots, where the arrow indicates direction of time.

METHOD

Participants

Participants were Android smartphone users recruited to participate in focus groups on media use in New York City, Chicago, and Los Angeles. At the completion of the focus group session, participants were approached about participating in a second study on laptop and smartphone media use. This analysis makes use of passively collected data obtained from N = 60 adults (34 female, 26 male) who were between age 20 and 50 years (M_{Age} = 34.77, SD_{Age} = 8.81). Total observation spanned 3,055 hours of adults' normal smartphone media use (2,199,654 screenshots), of which 121 hours of use (87,182 screenshots, 4.0%) were manually labeled for task switching behavior.

Screenshot (screenome) collection

The data collection procedure uses a proprietary software to capture screenshots at five-second intervals during device use, store them on local devices, and encrypt and transmit bundles of screenshots to research servers at intervals that accommodate constraints in bandwidth and device memory [11]. The data, sequences of screenshots we refer to as a "screenome", are collected in a fully unobtrusive manner that requires no user input, and thus minimizes both user and researcher costs often associated with behavioral observation and experience sampling. The screenshot data consist of image (jpg) files that contain a picture of whatever was visible on the user's smartphone screen (e.g., web browser, home-screen, e-mail program) at each moment.

Manual labeling of task switching

Randomly selected subsets of screenshots were manually labeled to obtain ground truth data on user behavior that could be used to train and evaluate machine learning algorithms that might be useful for labeling the remaining data. Manual labeling of big data often uses public crowd-sourcing platforms [2]. However, confidentiality and privacy protocols associated with screenome data require that labeling could only be done by members of the research team that are authorized to see the raw data. Here, labeling was done using *datavyu*, an opensource software widely used in manual coding of behavioral observation video [6]. Sequences of screenshots are presented in original time order, with each screenshot coded in a binary manner for whether it was a continuation of the prior task/behavior or a switch to a new task (e.g., from home screen to listening to news as shown in Fig. 2). Notable for the analysis, the time cost of labeling is substantial (human annotators need, on average, two hours to annotate 1000 screenshots).

Feature extraction

Visual stimulation. Extent of visual stimulation is related to individuals' judgements of appeal and engagement [16]. Screenshots of screen content were each converted from color to grayscale, with entropy then computed for across all pixels in the screenshot as a sum of the proportion of pixels with each grayscale color, weighted by its logarithm.

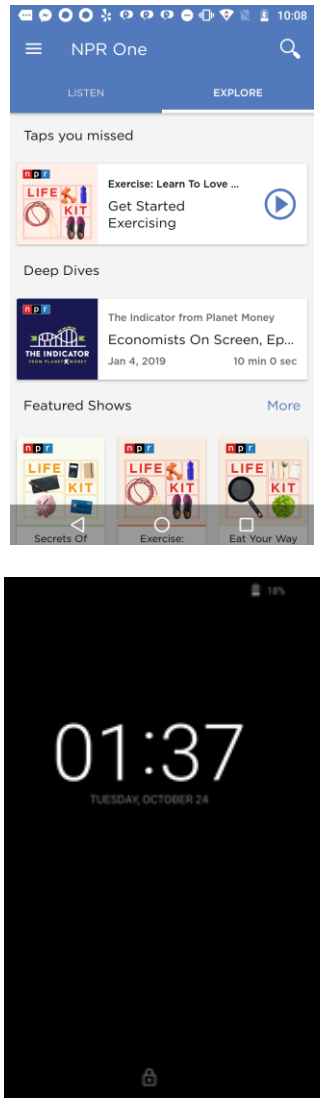

Figure 3: High versus low color complexity and word count (top: entropy = 6.2, word count = 33; bottom: entropy = 0.2, word count = 4)

As shown in Fig. 3, the screenshot in the top panel has higher entropy (entropy = 6.2) and is more visually stimulating than the relatively homogenous screenshot in the bottom panel (entropy = 0).

Cognitive load. The amount of symbolic information on the screen is sometimes considered a measure of cognitive load [4]. Previously, we have developed the facility to extract the text from each screenshot using a Tesseract-based Optical Character Recognition (OCR) module with character-level accuracy of 74% [3]. Word count on each screenshot is obtained by counting the number of strings separated by space in the OCR text. For example, for OCR text of a screenshot is "W Â« 41 B 7:49AM BREAKING: Double Palm Sunday Church", the word count is 10.

Velocity and accumulation. Previous literature suggests that the motivation processing system is influenced by not only the characteristics of the media content (e.g., valence), but also how quickly the content changes (velocity, acceleration) and accumulates [13]. Here, these features are proxied using the entropy and word count of the prior two screenshots (lag-1 and lag-2 as complements of first and second derivatives). In line with the limited capacity model [7], accumulation is calculated as the sum of entropy and sum of word count across all screenshots since the beginning of a smartphone use session (i.e., integration across time).

Sentiment. Sentiment of media content is known to influence individuals' attentional focus, emotionality, and thinking styles [12] – i.e., differential invocation of appetitive and aversive motivations. Here, the text extracted from each screenshot [9] was quantified using a Linguistic Inquiry of Word Count (LIWC [8]) approach wherein percentage of words that fall in various psychologically meaningful categories, such as positive and negative emotions, sociability, pronouns, etc. are quantified using 93 variables.

Time-related features. Circadian patterns influence individuals' subjective alertness and cognitive performance [5], which might influence task switching behavior – i.e., when people are more alert and have better cognitive performance during particular hours of each day, they may be more likely to switch to different tasks. Therefore, we include time-of-the-day in the time-related features. We also include time gap from the previous session as a time-related feature to explore the effect of lack of smartphone use on switching behavior; similarly, time since session-onset is included as a feature to examine the effect of elapsed time of smartphone use.

Classification

Our goal is to predict the switch points in the stream of screenshots, and we used a supervised machine learning method, random forest [1], with all of the above features. In brief, this tree-based method iteratively partitions the multidimensional feature space to identify the optimal splits in each feature that optimize classification of each screenshot as a switch or non-switch from the prior task or application. The ensemble of trees together facilitate possibility of identifying complex, nonlinear and multivariate relations between features and task switching, and evaluation of the importance of each feature in making accurate predictions of task switching.

Table 1: 10 Most Important Variables Based on Random Forest

Feature	*Importance*
Entropy velocity	34.8
Entropy	27.2
Entropy acceleration	26.3
Lag1 entropy	25.8
Word acceleration	16.9
Word velocity	16.9
Lag2 entropy	16.0
Lag1 word count	15.2
Accumulated entropy	15.2
Accumulated word count	13.7

ACKNOWLEDGMENTS

This work was partially supported by the National Institutes of Health (R01 HD076994, P2C HD041025, UL1 TR002014, T32 AG049676), the National Science Foundation I/UCRC Center for Healthcare Organization Transformation (CHOT, NSF I/UCRC award #1624727), the Penn State Social Science Research Institute, the Stanford University Cyber Initiative, the Knight Foundation, and the Stanford Child Health Research Institute, and the Stanford University PHIND (Precision Health and Integrated Diagnostics) Center

RESULTS AND DISCUSSION

Accuracy

The behavior of interest, task switching, is a relatively rare event (i.e., unbalanced binary outcome). At the sample level, only 12.6% of screenshots are indicative of a task-switch. Thus, accuracy was evaluated using precision-recall [9]. The area under curve (AUC) of the precision-recall curve of the best random forest prediction model with 108 features was 0.77 with out-of-bag (OOB) error rate of 7.3%.

Variable importance

Relative importance of each of the top 10 features for prediction of task switching is shown in Table 1. The variable importance of the velocity and accumulation features ranged from 12.9 to 34.8. Among these variables, visual stimulation and its velocity and accumulation played an important role in task switching prediction, e.g., color entropy, and its first and second derivative are the top 3 most important variables. Sentiment and time-related features were relatively unimportant in prediction of task switching. The variable importance of sentiment features ranged from 0.4 to 11 (e.g., positive emotion = 3.3, negative emotion = 3.6), and time-related features ranged from 1.6 to 13.1 (e.g., hour of the day = 8.8).

In sum, in line with data that the smartphone user experience is becoming increasingly visual, especially among the younger generation [10], we find that the visual features are particularly important in prediction of task switching and that the fast-paced decision of task switching is more heavily dependent on the visual features of media content than on textual features (e.g., importance of entropy is greater than that of word count with importance of 10.7, which was not in the top 10 features).

CONCLUSIONS

In this paper, we explore the idea of using unobtrusive experience sampling to collect screenshots and demonstrating it is a viable option to use features of the screenshots to predict task switching behavior. Based on our findings, features that gauge visual stimulation, cognitive load, and the accumulation and velocity of these features are the most important features in the prediction, compared with sentiment and time-related features. Classification methods were effective in determining the most relevant features in task switching, and explore nonlinear relations and interactions among features. Future research can further analyze task switching patterns (e.g., switching frequency, switching time distribution) on smartphones based on the result of this study.

REFERENCES

[1] Leo Breiman, 2001. Random Forests. *Machine Learning, 45 (1),* 5–32. https://doi.org/10.1023/A:1010933404324.

[2] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling, 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6(1),* 3-5. https://doi.org/10.1177/1745691610393980

[3] Agnese Chiatti, Xiao Yang, Miriam Brinberg, MJ Cho, Anupriya Gagneja, Nilam Ram, Byron Reeves, and C. Lee Giles, 2017. Text Extraction from Smartphone Screenshots to Archive in situ Media Behavior. *Proceedings of the Ninth International Conference on Knowledge Capture (K-CAP 2017*). ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3148011.3154468

[4] Krista E. DeLeeuw, and Richard E. Mayer, 2008. A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load. *Journal of Educational Psychology, 100(1),* 223-234.

[5] Derk-Jan Dijk, Jeanne F. Duffy, and Charles A. Czeisler, 1992. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *Journal of Sleep Research, 1(2),* 112-117. https://doi.org/10.2147/CPT.S32586

[6] Rick O. Gilmore, 2016. From big data to deep insight in developmental science. *WIREs Cognitive Science, 7(2),* 112-126. https://doi.org/10.1002/wcs.1379

[7] Annie Lang, 2000. The limited capacity model of mediated message processing. *Journal of Communication, 50,* 46-70. https://doi.org/10.1111/j.1460-2466.2000.tb02833.x

[8] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn, 2015. The development and psychometric properties of LIWC2015. Austin, TX: University of Texas at Austin.

[9] Vijay V. Raghavan, Gwang S. Jung, and Peter Bollmann, 1989. A Critical Investigation of Recall and Precision as Measures of Retrieval System Performance. *ACM Transactions on Information Systems, 7(3),* 205-229.

[10] Nilam Ram, Xiao Yang, Mu-Jung Cho, Miriam Brinberg, Fiona Muirhead, Byron Reeves, and Tom Robinson (under review). Teen screenomes: Describing and interpreting adolescents' day-to-day digital lives.

[11] Byron Reeves, Nilam Ram, Thomas N. Robinson, James J. Cummings, Lee Giles, Jennifer Pan, Agnese Chiatti, MJ Cho, Katie Roehrick, Xiao Yang, Anupriya Gagneja, Miriam Brinberg, Daniel Muise, Yingdan Lu, Mufan Luo, Andrew Fitzgerald and Leo Yeykelis (in press). Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them. *Human-Computer Interaction.*

[12] Yla R. Tausczik, and James W. Pennebaker, 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29(1),* 24-54. https://doi.org/10.1177/0261927X09351676

[13] Zheng Wang, Annie Lang, and Jerome R. Busemeyer, 2011. Motivational processing and choice behavior during television viewing: An integrative dynamic approach. *Journal of Communication, 61,* 71-93. https://doi.org/10.1111/j.1460-2466.2010.01527.x

[14] Leo Yeykelis, James J. Cummings, and Byron Reeves, 2014. Multitasking on a single device; Arousal and the frequency, anticipation, and prediction of switching between media content on a computer. *Journal of Communication, 64,* 167-192. https://doi.org/10.1111/jcom.12070

[15] Leo Yeykelis, James, J. Cummings, and Byron Reeves, 2018.The fragmentation of work, entertainment, e-mail, and news on a personal computer: Motivational predictors of switching between media content. *Media Psychology, 21(3),* 377-402.

[16] Xianjun Sam Zheng, Ishani Chakraborty, James Jeng-Weei Lin, and Robert Rauschenberger. 2009. Correlating Low-Level Image Statistics with Users' Rapid Aesthetic and Affective Judgements of Web Pages. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Pages 1-10 (CHI 2009).* https://doi.org/10.1145/1518701.1518703