Exploring the Relationship Between Momentary Emotions & Smartphone Content: A Pilot Study

Katherine Roehrick
Stanford University
Stanford CA 94305, USA
kroehr@stanford.edu

Gabriella Harari
Stanford University
Stanford CA 94305, USA
gharari@stanford.edu

CHI 2020 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.XXXXXXX
*update the above block & DOI per your rightsreview confirmation (provided after acceptance)

Abstract

Using a novel dataset of 28,905 screenshots of actual smartphone activity, we explored the relationship between momentary emotions and on-screen, smartphone content as seen and produced by participants in the wild. 13 participants installed the *Stanford Screenomics* application (a custom-built smartphone app that continually collects screenshots of smartphone activity) and completed experience sampling reports of their momentary emotions up to five times per day during a two-week study. We examined smartphone content by analyzing the on-screen text that participants both passively consumed and actively produced. Using the Linguistic Inquiry and Word Count (LIWC) program to perform text analyses, we found that smartphone content (e.g., on-screen text with higher amounts of cognitive or emotional words) was associated with momentary emotional experience (e.g., annoyance, boredom, and sadness). While these findings must be replicated in a larger study, they indicate that studying actual smartphone content can reveal new insights into the relationships between people's emotional experiences and their digital activity. We suggest that future research may eventually use passively collected smartphone content in order to create personalized models that predict

momentary emotions, thereby enabling unobtrusive observation of emotions in daily life.

Author Keywords

Emotions; language; experience sampling; naturalistic observation, smartphones; screenomics; screenome.

CSS Concepts

• Human-centered computing~Empirical studies in ubiquitous and mobile computing

Introduction

In order to study the relationship between personal behavior and emotional experience, many researchers have capitalized on the rich digital footprint that people leave behind with their mobile phones [5, 10]. For example, the mobile sensing research community has shown that many of the behaviors, such as physical activity, sleeping patterns, and social interaction, that correlate with emotional experience can be captured through smartphone sensor data [15]; for example, smartphone sensing can be used to measure physical activity via the accelerometer [12], mobility patterns via GPS data [20, 21], sleeping patterns via the light sensor [1], and social interactions via phone logs [16].

In addition to sensor data, smartphones can provide a unique record of the content that people both actively create and passively consume. Notably, previous research has consistently found that both one's active creation and passive consumption of content can reveal their cognitive and emotional states [e.g., 4, 23]. For example, scholars have used social media posts and music streaming data to predict diurnal and seasonal mood patterns [8, 17].

We build on this rich body of mobile sensing and social media literature by studying screenshots of actual smartphone content as seen by participants in the wild. While researchers have consistently found that "what people read and produce could affect what people do and feel" [23], we are among the first—if not the only—researchers to use actual smartphone content obtained in the wild and continually throughout the day in order to report on the dynamic relationship between momentary emotional experience and digital activity. Future research may eventually use passively collected smartphone content in order to unobtrusively measure momentary emotions in daily life through personalized models that predict emotional experience.

Method

Participants and Procedures

Data collection and processing procedures were approved by the Institutional Review Board at Stanford University. The present analyses makes use of data collected from 13 students (9 women, 4 men) recruited from two Northern California colleges for class credit. Participants ranged in age from 18 to 56 (median = 26) and identified as Asian/Pacific Islander (n = 6), Hispanic/Latino (n = 5), Caucasian (n = 1), or multiple ethnicities (n = 1).

Participants enrolled in the study online, where they provided consent and were given instructions on how to install the custom-built *Stanford Screenomics* application onto their Android smartphone. The application captured timestamped screenshots every five seconds during smartphone use. The screenshots were securely encrypted and transmitted every 24 hours to research servers, which are accessible only to research staff trained and certified for conducting human subjects research. Screenshot collection lasted for up to two consecutive weeks. At the end of the study, participants were thanked and reminded to uninstall the *Stanford Screenomics* application.

In addition to screenshot data, participants provided self-reported surveys regarding their current emotional state up to five times per day during their participation. Only participants who provided screenshots and self-reported surveys for more than three days are included in this analysis.

Ethical and Privacy Considerations

This project has collected an unprecedented amount of data about individuals' smartphone use. While we believe that this data is necessary to understand the dynamic nature of emotional experience and its relationship to smartphone use, we are committed to reducing the risks, such as de-anonymization or information leakage, that hyper-personal, cross-contextual data and smartphone analytics pose for participant privacy. In collaboration with the Stanford IRB, we developed a system to ensure proper handling of sensitive data and respect for participant privacy. Specifically, we collected data using a rigorous privacy protocol that restricts data access to a small set of university-based researchers for academic study only. Unlike many corporate data-use agreements, our privacy protocol does not allow data to be used for any other purposes or by any other parties.

Experience Sampling Procedures and Measures

During the two-week study, participants were texted a survey link at random hours five times each day. In the survey, participants were asked to rate (0 = *Not At All*, 4 = *Very Much*) how they felt during the past 15 minutes using a modified version of the 10-item D-FAW, which includes measures assessing anxiety, anger, displeasure, tiredness, and boredom [19]. Unlike the short PANAS, one of the widely used measures of emotional well-being, the D-FAW captures low activation terms for both positive and negative affect [19]. In addition to the D-FAW, participants were asked to rate (0 = *Not At All*, 4 = *Very Much*) how lonely they felt during the past 15 minutes. Only participants who completed at least one survey for three days or more are included in this analysis.

Screenshot Processing Procedures and Measures

Screenshots were processed to obtain a record of all the text that appeared on each screen (see processing details in [2]). In brief, each screenshot was first converted to grayscale and then binarized into white and black pixels to separate the textual foreground (white pixels) from the surrounding background (black pixels). After segmenting, each candidate block of text was then passed to a Tesseract-based optical character recognition (OCR) text extraction module in order to compile Unicode text files for each screenshot. Compared to ground-truth transcription of a random subset of images, the implemented OCR method achieved 74% accuracy at the individual character level [2].

After processing, we matched the self-reported surveys to the screenshots that were captured up to 15 minutes prior to the participants starting the self-reports (matched screenshots = 28,905). We then averaged the language features, described below, within each 15-minute screenshot sequence to obtain a set of aggregate language features for each matched survey report. This process resulted in a total of 549 matched reports.

Language features of the extracted text was quantified using the LIWC (Linguistic Inquiry and Word Count) dictionary [18]. In brief, LIWC is a dictionary-based method used to examine the psychological content of both personal speech and writing, and has been used to study content in diverse kinds of documents, including diaries, blog posts, lyrics, poems, and tweets (e.g, [4, 22]). Here, we used LIWC to quantify the percentage of words in each screenshot classified as emotional, social, profane, and cognitive. Cognitive words, which are represented by causal or insight-oriented words, have been associated with actively thinking about and coming to terms with threatening or stressful events [3].

Active Use (i.e., when a participant is actively producing content) was ascertained by identifying which screenshots contained a digital keyboard, recognized by detection of a QWERTY string within the screenshot. Screenshots without a QWERTY string were labeled as passive use (i.e., when a participant is passively consuming content).

Data Analysis

We examined the relationship between smartphone content and emotion using a series of multilevel models predicting momentary emotions from the language features (Level 1) nested within persons (Level 2). We used the R library ordinal (Christensen 2019) to conduct multilevel models for ordinal categorical data (Christensen 2019). All of the previously mentioned LIWC categories (emotional, social, profane, and cognitive) were entered into the models as person mean centered covariates. The models used random intercepts at Level 2. We compared a baseline, intercepts-only model to models with the language features added as predictors. Because much of the previous literature predicting emotion from digital content uses social media posts, or "actively produced" content, we analyzed models exploring 1) all screenshots and 2) only those screenshots labelled as "active use".

Results

First, we analyzed model fit for all screenshots (i.e., both active and passive use combined). Compared to baseline, we found that adding language features as covariates significantly improved model fit when predicting annoyance ($\chi^2 = 0.029$). Upon examination of the fixed effects model, we found that participants were less likely to report feeling annoyed as emotional screen content increased ($\beta = -0.153$, $p = 0.0472$). In other words, for each percentage increase in the emotional content seen on screen, the odds of participants reporting a higher level of annoyance declined by 14% ($e^{-0.153} = 0.86$).

Next, we analyzed model fit for active use screenshots alone. We found that adding language features as covariates significantly improved model fit when predicting reports of boredom ($\chi^2 = 0.037$). Likewise, the model with language covariates is nearly significantly better than baseline when predicting reports of sadness ($\chi^2 = 0.056$). Upon examination of these models, it appears that individuals were less likely to report being bored when they were using cognitive words as part of their smartphone use ($\beta = -0.417$, $p = 0.009$). In other words, for each percentage increase in cognitive words used, the odds of participants reporting a higher level of boredom declined by 34% ($e^{-0.417} = 0.66$). This suggests that people are less likely to be bored the more they are engaged in complex thinking while using their phone. On the other hand, participants are also more likely to report being sad when using cognitive words on their phone ($\beta = 0.430$, $p = 0.026$). That is to say, for each percentage increase in cognitive words, the odds of participants reporting a higher level of sadness increased by 54% ($e^{0.430} = 1.54$). As previously mentioned, cognitive words have been associated with actively thinking about stressful or threatening events as a means to move on. With this in mind, this result suggests that participants may be using their smartphone during these times in ways that provide them with a means of processing their momentary sadness.

Discussion

Many researchers have capitalized on the rich digital footprint that people leave behind with their mobile phones as a means to study the relationship between personal behavior and emotional experience. In contrast to prior studies, this study is the first to use actual smartphone screen content obtained in the wild and throughout the day in order to explore how on-screen language is related to momentary emotions. Compared to studies using social media language, which rely on sporadically created posts, our study examines the full range of content that people both see and produce on their smartphones continually throughout the day. Importantly, we find that by separately examining 1) all screen content and 2) screen content captured during active use, different insights into people's emotional experience can be ascertained. While our findings must be replicated in a larger study, they indicate that studying actual smartphone content can reveal new insights into the relationships between people's emotional experiences and their digital activity. Notably, if we can link screen content to momentary emotional experience, we may be able to predict what people are feeling in real-time. We therefore suggest that smartphone content may enable unobtrusive observation of momentary emotion as experienced in daily life.

Limitations

This paper presents exploratory findings from a pilot study regarding the relationship between screen content and momentary emotional experience. As such, the generalizability of the findings to the general population is limited by the small size and the unique characteristics of the participants (i.e., predominantly young adults obtaining higher education). Moreover, because this initial research exploration uses a small sample size (N = 13), we have low statistical power; therefore, the results which we found are encouraging. As the dataset grows, we will gain more power to test and develop new research questions about the dynamic relationship between digital activity and momentary emotion.

Second, despite having a near exact record of screen content, we only extracted a relatively small set of language features from each screenshot. Future studies could make use of a more expanded feature set, including image analysis, in order to further understand the dynamic relationship between screen content and momentary emotional experience. Relatedly, in order to further understand this relationship, additional analytical methods should be explored. For example, neural networks, which have previously been used to study emotion recognition in images and speech, may be a valuable tool in analyzing our feature-rich dataset [e.g., 6, 9, 11, 24]. Additionally, we plan to further explore the intraindividual dynamics of digital activity and emotional experience in future studies with idiographic network-based analyses that use contemporaneous and lagged associations, a method which has previously been used successfully to study intraindividual variation in psychological wellbeing [7].

Notwithstanding these limitations, our results illustrate the potential of passively-collected smartphone content to not only generate new insights about the dynamic relationship between smartphone use and momentary emotions but also predict momentary emotions as they occur in daily life.

Acknowledgments

The authors would like to acknowledge The Stanford Institute for Human-Centered Artificial Intelligence and The Stanford Screenomics Lab for supporting this project. Finally, the authors would like to thank Melvin and Joan Lane for sponsoring this work through the Stanford Graduate Fellowship Program.

References

[1] Chen, Zhenyu, Mu Lin, Fanglin Chen, Nicholas D. Lane, Giuseppe Cardone, Rui Wang, Tianxing Li, Yiqiang Chen, Tanzeem Choudhury, and Andrew T. Campbell. "Unobtrusive sleep monitoring using smartphones." In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 145-152. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2013.

[2] Chiatti, Agnese, Mu Jung Cho, Anupriya Gagneja, Xiao Yang, Miriam Brinberg, Katie Roehrick, Sagnik Ray Choudhury, Nilam Ram, Byron Reeves, and C. Lee Giles. "Text extraction and retrieval from smartphone screenshots: Building a repository for life in media." In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pp. 948-955. ACM, 2018.

[3] Creswell, J. David, Suman Lam, Annette L. Stanton, Shelley E. Taylor, Julienne E. Bower, and David K. Sherman. "Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing?." *Personality and Social Psychology Bulletin* 33, no. 2 (2007): 238-250.

[4] De Choudhury, Munmun, Michael Gamon, Scott Counts, and Eric Horvitz. "Predicting depression via social media." In *Seventh international AAAI conference on weblogs and social media*. 2013.

[5] DeMasi, Orianna, Sidney Feygin, Aluma Dembo, Adrian Aguilera, and Benjamin Recht. "Well-being tracking via smartphone-measured activity and sleep: cohort study." *JMIR mHealth and uHealth* 5, no. 10 (2017): e137.

[6] Ebrahimi Kahou, S., Michalski, V., Konda, K., Memisevic, R., & Pal, C. (2015, November). Recurrent neural networks for emotion recognition in video. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (pp. 467-474).

[7] Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of abnormal psychology*, *126*(8), 1044.

[8] Golder, Scott A., and Michael W. Macy. "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures." *Science* 333, no. 6051 (2011): 1878-1881.

[9] Han, K., Yu, D., & Tashev, I. (2014). Speech emotion recognition using deep neural network and extreme learning machine. In *Fifteenth annual conference of the international speech communication association*.

[10] Harari, Gabriella M., Sandrine R. Müller, Min SH Aung, and Peter J. Rentfrow. "Smartphone sensing methods for studying behavior in everyday life." *Current opinion in behavioral sciences* 18 (2017): 83-90.

[11] Kim, H. R., Kim, Y. S., Kim, S. J., & Lee, I. K. (2018). Building emotional machines: Recognizing image emotions through deep neural networks. *IEEE Transactions on Multimedia*, *20*(11), 2980-2992.

[12] Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore. "Activity recognition using cell phone accelerometers." *ACM SigKDD Explorations Newsletter* 12, no. 2 (2011): 74-82.

[13] Li, Y., Chen, F., Li, T. J. J., Guo, Y., Huang, G., Fredrikson, M., ... & Hong, J. I. (2017). PrivacyStreams: Enabling Transparency in Personal Data Processing for Mobile Apps. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 76.

[14] Lindquist, Kristen A., Jennifer K. MacCormack, and Holly Shablack. "The role of language in emotion: Predictions from psychological constructionism." *Frontiers in Psychology* 6 (2015): 444.

[15] Low, C. A., Dey, A. K., Ferreira, D., Kamarck, T., Sun, W., Bae, S., & Doryab, A. (2017). Estimation of symptom severity during chemotherapy from passively sensed data: exploratory study. *Journal of medical Internet research*, *19*(12), e420.

[16] Min, Jun-Ki, Jason Wiese, Jason I. Hong, and John Zimmerman. "Mining smartphone data to classify life-facets of social relationships." In Proceedings of the 2013 conference on Computer supported cooperative work, pp. 285-294. ACM, 2013.

[17] Park, Minsu, Jennifer Thom, Sarah Mennicken, Henriette Cramer, and Michael Macy. "Global music streaming data reveal diurnal and seasonal patterns of affective preference." *Nature Human Behaviour* 3, no. 3 (2019): 230.

[18] Pennebaker, James W., Martha E. Francis, and Roger J. Booth. "Linguistic inquiry and word count: LIWC 2001." *Mahway: Lawrence Erlbaum Associates* 71, no. 2001 (2001): 2001.

[19] Russell, Emma, and Kevin Daniels. "Measuring affective well-being at work using short-form scales: Implications for affective structures and participant instructions." *human relations* 71, no. 11 (2018): 1478-1507.

[20] Saeb, S., Lattie, E. G., Kording, K. P., & Mohr, D. C. (2017). Mobile Phone Detection of Semantic Location and Its Relationship to Depression and Anxiety. JMIR mHealth and uHealth, 5(8). doi:10.2196/mhealth.7297 [208 participants]

[21] Saeb, Sohrab, Emily G. Lattie, Stephen M. Schueller, Konrad P. Kording, and David C. Mohr. "The relationship between mobile phone location sensor data and depressive symptom severity." *PeerJ* 4 (2016): e2537.

[22] Tausczik, Yla R., and James W. Pennebaker. "The psychological meaning of words: LIWC and computerized text analysis methods." *Journal of*

language and social psychology 29, no. 1 (2010): 24-54.
[23] Wang, N., Michal Kosinski, D. J. Stillwell, and J. Rust. "Can well-being be measured using Facebook status updates? Validation of Facebook's Gross National Happiness Index." *Social Indicators Research* 115, no. 1 (2014): 483-491.
[24] Zheng, W. Q., Yu, J. S., & Zou, Y. X. (2015, September). An experimental study of speech emotion recognition based on deep convolutional neural networks. In *2015 international conference on affective computing and intelligent interaction (ACII)* (pp. 827-831). IEEE.