Killing-Time Detection from Smartphone Screenshots

Yu-Chun Chen*
National Yang Ming Chiao Tung University
lesley.yc.chen@gmail.com

Keui-Chun Kao*
National Yang Ming Chiao Tung University
johnson0213.cs07@nycu.edu.tw

Yu-Jen Lee
National Yang Ming Chiao Tung University
lee.cs09@nycu.edu.tw

Faye Shih
Bryn Mawr College
fshih@brynmawr.edu

Wei-Chen Chiu
National Yang Ming Chiao Tung University
walon@cs.nctu.edu.tw

Yung-Ju Chang
National Yang Ming Chiao Tung University
armuro@cs.nctu.edu.tw

ABSTRACT

Finding good moments to deliver interruptions has drawn research attention. Since users have attention surplus at these moments, killing-time is considered one such a kind of moment. However, detection on killing-time has been under researched. In this paper, we propose a screenshot-based killing-time detection with deep learning. Our model achieves an accuracy 79.71%, recall 90.24%, precision 84.51%, and AUROC 65.50%. This suggests that using screenshots to detect users' kill time behavior on smartphones is a promising approach. It may be worthwhile to investigate how the fusion of screenshots and sensor data can further improve detection.

CCS CONCEPTS

• **Human-centered computing → Smartphones**; **Ubiquitous and mobile computing systems and tools**.

KEYWORDS

Kill time; Screenshot; Deep Learning; Opportune Moment

ACM Reference Format:
Yu-Chun Chen, Keui-Chun Kao, Yu-Jen Lee, Faye Shih, Wei-Chen Chiu, and Yung-Ju Chang. 2021. Killing-Time Detection from Smartphone Screenshots. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp-ISWC '21 Adjunct), September 21–26, 2021, Virtual, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3460418.3479295

1 INTRODUCTION

The identification of a good moment to deliver interruptions, such as notifications [9], questionnaires [9], reading material [6, 9], or crowdsourcing tasks [4] on the phone, has drawn many calls for research attention. The majority of these works intended to protect users' attention by delivering content when users are interruptible. In another line of work, researchers are intended to find when users have attention surplus and sought to leverage that attention, such as when they are bored [10] or is waiting [3, 8], to deliver digital content. Machine learning techniques have been commonly adopted for detecting these moments (e.g., [9, 10], in which information such as phone sensors, phone status, users' recent actions on the phone have been found to show strong indications of users' attention on the phone [1]). However, they are yet found not as effective in predicting moments of attention surplus [10]. One possible explanation is that checking notifications is a pervasive and intermittent activity on the phone [5]; thus being able to spend quick attention on checking notifications does not necessarily indicate attention surplus. Another one is that boredom is an unobservable state of mind, whereas killing time is an observable behavioral outcome due to attention surplus, a common activity users conduct when attempting to filling perceived free time [2]. Therefore, we deem that killing-time behavior may display patterns that can be observed from users' phone activities, which may differ from when s/he is not killing time. Hence, we propose screenshot-based detection–using phone screenshots, which contains rich information about users' phone activity, to detect killing-time using both convolutional neural network and recurrent neural network. Our preliminary results are based on a dataset consisting of 215,807 screenshots from six users, who recorded and labeled their phone screenshots over 14 days. Our model achieves an accuracy 79.71%, recall 90.24%, precision 84.51%, F1 86.88%, and AUROC 65.50%. Our future work will seek to reduce false-positive rates, fuse sensor data and screenshots to improve the performance, collect more users' data, and examine the model's generalizability across users.

2 METHODOLOGY

2.1 Data Collection

We develop an Android app that automatically collects screenshots and phone sensor data every 5 seconds. We design a user interface for users to easily select a group of screenshots via drag-and-drop for data labeling. Four possible labels are available for labeling that, at the moment when screenshots were taken, 1) whether he/she was killing time and 2) whether he/she was available for viewing notifications. Users have the authority to decide whether to upload the screenshots to protect their privacy.

Our dataset in total consists of 215,807 screenshots from six users (3 females and 3 males; 22-32 years old; 3 students and 3 employed) recruited via social media. 77.7% of the screenshots are labelled as "killing time". All users participated in data collection for fourteen days, and were paid NT$1200 ($43.34 USD).

*Both authors contributed equally to this research.

UbiComp-ISWC '21 Adjunct, September 21–26, 2021, Virtual, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8461-2/21/09.
https://doi.org/10.1145/3460418.3479295

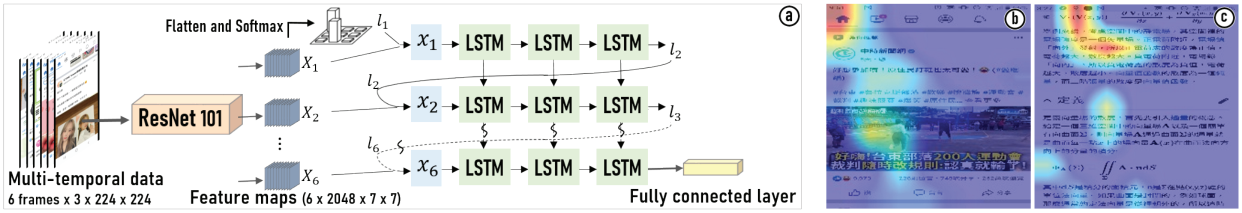

Figure 1: (a) Architecture overview of our proposed model for predicting user's killing time from consecutive screenshots. Visualizing attention map extracted from last convolution layer of (b) killing time, and (c) not killing time.

2.2 Model Description

We build up a deep neural network (see Fig. 1 (a)) to predict the user's killing-time from 6 temporally consecutive screenshots, which are resized into 224×224 pixels. Our network first extracts the feature map of size 7×7×2048 from each screenshot via a ImageNet-pretrained Resnet-101 [7] backbone, where the feature maps are sequentially taken as input for a 3-layer long short-term memory (LSTM) model to reach the final output of killing-time detection. The dimensionalities of the hidden state, cell state, and the hidden layer of our LSTM model are set to 512, 512, and 256, respectively. In particular, when the feature map X_t at time t feed into the LSTM, it is spatially weighted according to the attention map l_t predicted from the previous time step $t-1$ in order to better account for the regions on the screenshot that could be informative to detecting the killing-time activity. For the feature map X_1 of the first screenshot, we adopt the self-attention mechanism to obtain the corresponding attention map l_1 (denoted as "flatten and softmax" in Fig. 1 (a)). Fig. 1 (b) and (c) are two examples of visualized attention

3 EXPERIMENT

The dataset is split into two parts — the training (80%) and the test (20%). We utilize the Adam optimizer to train our model for 20 epochs with weight decay penalty set to 10^{-3} and adopt 5-fold cross-validation to produce the experimental results (see Table 1). Despite the small size of the dataset, the results show promising results. The model achieves an accuracy of 79.71%, recall 90.24%, precision 84.51%, F1 86.88% and AUROC 65.50%. The performance is advantageous by only using screenshot data to detect boredom state [10], which it achieves both high recall and precision without needing to trade off either one. On the other hand, the false-positive rate, i.e. mis-recognizing not-killing-time moment as killing-time moment, is still high (59%). Thus, our future work will seek to lower the false-positive rate; meanwhile, we will gain more users' screenshots and add fusion of sensor and screenshot data to further improve the model.

4 CONCLUSION

We employed deep learning to detect when users are killing time on smartphones using their screenshots. In a two-weeks study with six participants, the model achieves an accuracy of 79.71%, recall 90.24%, precision 84.51%, F1 86.88% and AUROC 65.50%. These results show that using screenshots for detecting users killing time is promising. We will further improve the model, including add fusion of sensor data and screenshots.

Table 1: Confusion Matrix

	not killing time	killing time
not killing time	0.41	0.59
killing time	0.10	0.90

🟦 Predicted
🟩 Actual

ACKNOWLEDGMENTS

We thank Ministry of Science and Technology, R.O.C for their support (107-2218-E-009-030-MY3, 109-2221-E-009-097). We also thank National Center for High-performance Computing, R.O.C for providing computational and storage resources.

REFERENCES

[1] Christoph Anderson, Isabel Hübener, Ann-Kathrin Seipp, Sandra Ohly, Klaus David, and Veljko Pejovic. 2018. A survey of attention management systems in ubiquitous computing environments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 1–27.

[2] Barry Brown, Moira McGregor, and Donald McMillan. 2014. 100 Days of IPhone Use: Understanding the Details of Mobile Device Use. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services* (Toronto, ON, Canada) *(MobileHCI '14)*. Association for Computing Machinery, New York, NY, USA, 223–232. https://doi.org/10.1145/2628363.2628377

[3] Carrie J. Cai, Philip J. Guo, James R. Glass, and Robert C. Miller. 2015. *Wait-Learning: Leveraging Wait Time for Second Language Education.* Association for Computing Machinery, New York, NY, USA, 3701–3710. https://doi.org/10.1145/2702123.2702267

[4] Chia-En Chiang, Yu-Chun Chen, Fang-Yu Lin, Felicia Feng, Hao-An Wu, Hao-Ping Lee, Chang-Hsuan Yang, and Yung-Ju Chang. 2021. "I Got Some Free Time": Investigating Task-execution and Task-effort Metrics in Mobile Crowdsourcing Tasks. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3411764.3445477

[5] Tilman Dingler and Martin Pielot. 2015. I'll Be There for You: Quantifying Attentiveness towards Mobile Messaging. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/2785830.2785840

[6] Tilman Dingler, Benjamin Tag, Sabrina Lehrer, and Albrecht Schmidt. 2018. Reading Scheduler: Proactive Recommendations to Help Users Cope with Their Daily Reading Volume. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (Cairo, Egypt) *(MUM 2018)*. Association for Computing Machinery, New York, NY, USA, 239–244. https://doi.org/10.1145/3282894.3282917

[7] Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* cs.CV (2016), 770–778.

[8] Ellen Isaacs, Nicholas Yee, Diane J Schiano, Nathan Good, Nicolas Ducheneaut, and Victoria Bellotti. 2009. Mobile Microwaiting Moments: The Role of Context in Receptivity to Content While on the Go.

[9] Martin Pielot, Bruno Cardoso, Kleomenis Katevas, Joan Serrà, Aleksandar Matic, and Nuria Oliver. 2017. Beyond Interruptibility: Predicting Opportune Moments to Engage Mobile Phone Users. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 91 (Sept. 2017), 25 pages. https://doi.org/10.1145/3130956

[10] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. Association for Computing Machinery, New York, NY, USA, 825–836. https://doi.org/10.1145/2750858.2804252