Capturing and Predicting User Frustration to Support a Smart Operating System

Sepideh Goodarzy
Eric Keller
Maziyar Nazari
Eric Rozner
sepideh.goodarzy@colorado.edu
eric.keller@colorado.edu
maziyar.nazari@colorado.edu
eric.rozner@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

Richard Han
Mark Dras
Young Lee
Deborah Richards
richard.han@mq.edu.au
mark.dras@mq.edu.au
young.lee@mq.edu.au
deborah.richards@mq.edu.au
Macquarie University
Sydney, NSW, Australia

ABSTRACT

This paper presents an IRB-approved human study to capture data to build models for human frustration prediction of computer users. First, an application was developed that ran in the user's computer/laptop/VM with Linux 20.04. Then, the application collected a variety of data from their computers, including: mouse clicks, movements and scrolls; the pattern of keyboard keys clicks; user audio features; and head movements through the user video; System-wide information such as computation, memory usage, network bandwidth, and input/output bandwidth of the running applications in the computer and user frustrations. Finally, the application sent the data to the cloud. After two weeks of data collection, supervised and semi-supervised models were developed offline to predict user frustration with the computer using the collected data. A semi-supervised model using a generative adversarial network (GAN) resulted in the highest accuracy of 90%.

CCS CONCEPTS

• **Software and its engineering → Operating systems**; • **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Computer-based User Frustration, Intelligent Operating System

ACM Reference Format:
Sepideh Goodarzy, Eric Keller, Maziyar Nazari, Eric Rozner, Richard Han, Mark Dras, Young Lee, and Deborah Richards. 2023. Capturing and Predicting User Frustration to Support a Smart Operating System. In *28th International Conference on Intelligent User Interfaces (IUI '23 Companion), March 27–31, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3581754.3584124

IUI '23 Companion, March 27–31, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0107-8/23/03.
https://doi.org/10.1145/3581754.3584124

1 INTRODUCTION

Technology frustration [1], such as computer-based frustration experienced by early computer users in the workplace [18] due to slow response times and system crashes, continue to be experienced by computer users in a wide range of applications and contexts as diverse as smart home products [4] and advertiming [29]. These home and workplace frustrations can lead to stress and poor health, particularly since instead of the system adapting to the user, the user is required to adapt their work processes to compensate [14]. While current interest in dealing with frustration can be found within the HCI community, most HCI research focuses on how to relieve frustration with specific computer applications. We believe there is also scope for reducing computer frustration at the OS level.

The main job of the operating system is to manage and distribute the computer resources among the applications. Ideally, the OS would be able to allocate resources based on what is of importance to the user. However, due to the challenge of determining what tasks have greatest current priority to an individual, current approaches are unable to adapt to specific user needs and typically assign the same static priority across all CPU processes such as the Linux Completely Fair Scheduler (CFS) [24]. Another approach is to statically prioritise tasks in the foreground on the assumption they are the most important, as in the recent Acclaim system [23], However, such an assumption does not hold for many common situations such as when one is waiting for an important download to complete or multitasking, e.g. listening to music while working. Even if one is able to identify when the user is frustrated, the source of the frustration, and indeed whether it is computer-related, may not be evident. The issue with current approaches is that they fail to obtain feedback from the user.

As a first step to create a smart operating system (SmartOS) that automatically learns user behavior and can allocate more resources to user priorities to alleviate their frustrations, this paper presents a study that captures and utilises user feedback about their frustration with the computer to build user models of frustration, an overview of the application and the collected data from 15 users, and a comparison between a supervised and semi-supervised prediction model [11].

2 RELATED LITERATURE

There are two main types of emotions: primary and secondary emotions. Primary emotions are the more impulsive ones, such as fear, and secondary ones go through cognitive procedures such as frustration, pride, and satisfaction. The secondary emotions are the main focus of Human-Computer Interaction (HCI) [2]. This brief reviews data used and methods used to predict emotion which have informed the user data collected in our study and our models.

To infer emotion [16, 34] studied the captured mouse cursor's x/y position and timestamp at a millisecond precision rate, total distance, and average cursor's speed and [36] analyzed the number of mouse clicks per minute, the average duration of mouse clicks (from the button-down to the button-up event), and the maximum, minimum and average mouse speeds. Time duration between the proposed combinations of events was used to predict the user's affective state in [30]. Inspired by such studies, we captured the x/y positions for mouse cursor, clicks and scrolls and the scrolls' amount in the x/y coordinate along with timestamps at a millisecond precision rates.

According to other studies, frustrated users may betray their agitations by typing faster or clustering specific keys together [17, 20, 36]. In [17] they studied features such as typing speed, total typing time, number of backspaces and idle time to recognize emotions from keyboard stroke pattern.

[30] studied collected a range of key press data and the indicators proposed in [8] to predict the user affective state. Following the footsteps of the previous work, our SmartOS app collected all the keys clicks and special keys clicks (such as ESC, delete, enter, space, etc.) with timestamps at a millisecond precision rates.

Users may voice their frustration in various ways, such as raising their voice or vocalizing more rapidly or frequently. [32] used audio spectrum for emotion recognition using deep neural networks, while [28] they made an emotion recognition system based on prosodic features (i.e., intensity, pitch, formant frequencies of sounds) combined with short-term perceptual features to classify emotions. In [33], they used spectrograms, the existence of certain words, and a phoneme segmentator to recognize emotions. In [31] they used the perceptual evaluation of audio quality (PEAQ) model to provide a mathematical model resembling the human auditory system. To show the manual acoustic emotional features more intuitively, [3] visualized several features, including chroma, zero crossing rate, MFCCs, energy, and spectral flux. In [13] they showed that the MFCC are the best features for recognition of emotional content in the audio. Thus SmartOS app collected the MFCCS of users' audio.

Frustrated users may move their heads in divergent directions quickly. In [32] they used video frames with deep neural networks to recognize emotions from head movements. SmartOS collected head positions which were computed using convolutional neural network face detection model applied on video frames recorded with the user's webcam.

The previous research on emotion prediction is typically focused to a single modality, while our research is on human frustration prediction of computer users using highly multi-modal models. For example, In [9, 27, 35] they focused on using mouse data to predict emotions, applying linear regression, support vector machines, and random forest as their machine learning models. In [10, 17], they predicted emotion from the keyboard strokes' patterns using the same data as this paper, and they used multi-layer perceptron and random forest as their machine learning models. In [7, 15, 19, 21, 22, 26] they focused on emotion prediction using speech. In [19] they used a pre-trained image classification network. In [7, 15] they employed a deep learning architecture and in [21, 22], they applied a domain adversarial neural network and unsupervised learning respectively. In [12, 25] they used head motions to predict emotion using feature engineering and support vector regression, linear regression, hidden Markov model, LSTM, CNN, and the fusion between LSTM and linear regression.

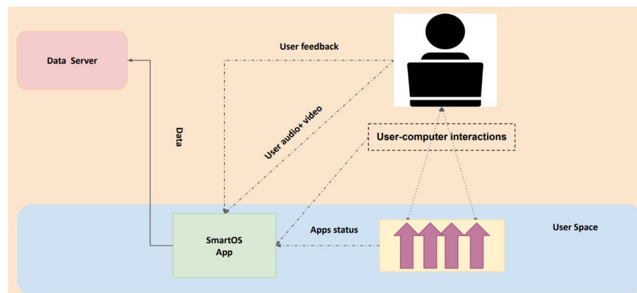

Figure 1: SmartOS App: Data Collection for Human Frustration Prediction Architecture

3 APPLICATION OVERVIEW

An IRB-approved study collected data from fifteen human users of desktop computers by installing a SmartOS app on their computers (Figure 1). including time, the list of running apps and their resource usage profiles and resource utilization of the system and the active app that the user was working with, mouse clicks and movements, keyboard clicks' patterns, audio features, head movements through computer's microphone and webcam and explicit feedback when users became frustrated. Linux as an open source OS provided easy access to OS system and application state. The collected data were used to train a model that predicts user frustration.

3.1 Application Usage Procedure and Data Collection

After giving consent, the user received an email link to install the SmartOS software app on their desktop/laptop. After an initial satisfaction "First Survey" in which the user describes the extent of their dissatisfaction with their computer and reasons for being dissatisfied, passive monitoring began on the user's desktop/laptop. Once the SmartOS server received the survey data, it generated a unique numerical ID for the user and registered it in the system for the remainder of the study. No personal data was collected. Audio data were converted on the user's computer to MFCCS (Mel-frequency cepstrum), then sent to the server. The webcam detected the user's head position with respect to their screen size using a convolutional neural network face detection model running on the user's computer, and it sent four numbers: top, bottom, left, and right of the user's face position with respect to their screen size to the server. As for the keyboard pressing patterns, the smartOS app only sent the timestamp the user pressed a key to the server.

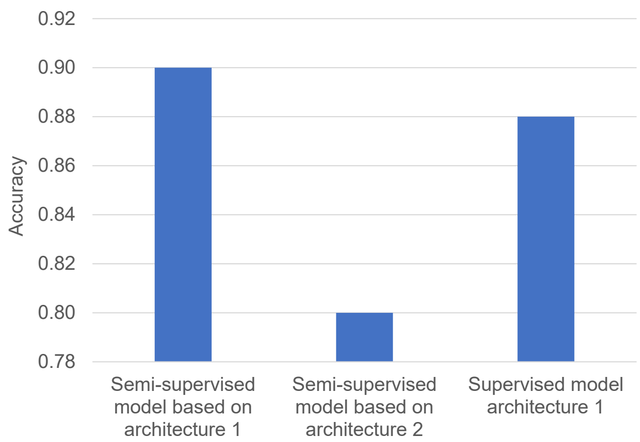

Figure 2: Human Frustration Prediction Models Comparisons.

The SmartOS app collected user data for two weeks. SmartOS' goal is to predict the cases when the user is frustrated with the system. The system resource allocation information helps identify scenarios where the user is specifically unhappy due to the slow response of their computer. Users reported their frustration with the system through using the Report Button, which opens a small window for describing and submitting the event causing the frustration to provide. Reports provided the ground truth value of user frustration with their computers.

4 EVALUATION

The human study was designed for data collection, so the SmartOS app did not modify the resource allocations policy on the user's computer. Due to the multi-modal input data, a supervised model using deep learning was defined, known as architecture 1. The supervised model is trained on the labeled data and a subset of the unlabeled data.

The numbers of labeled data were limited because users did not report all their frustration cases. Sometimes, they forgot to report their frustrations or became tired of reporting the same reason multiple times. In addition, users adjusted themselves and their expectations to avoid reporting some difficulties. Thus the unlabeled data should not be recognized as satisfaction cases. Hence, we modified the model's architecture to a semi-supervised model [5] joining a Generative Adversarial Network (GAN) [6] with generator, unsupervised and supervised models sharing base layers of architecture 1.

The evaluation was conducted iteratively over 256 epochs. The total amount of input data of the supervised model was 1330 cases. To balance the supervised model input data, half of the cases were the frustration cases, and the other half were randomly selected from unlabeled cases and interpreted as satisfaction cases. Given this balanced input data, accuracy was used as the primary evaluation metric for model comparison. The total number of data collected from our user study was 122577 cases that were fed to the semi-supervised model. The input data of the supervised model was divided into 80% training data and 20% test/validation data. The skit-learn train_test_split function did the sampling for the train and test data with the stratify parameter set according to the labels which results in unbiased test results. We also compared to an approach taken by prior work [12, 25] that used a more complicated second architecture for their semi-supervised model, which we deem architecture 2.

The semi-supervised model was compared with the supervised model. As shown in the graph of Figure 2 the semi-supervised model based on our architecture 1 reached the highest accuracy of 90%. The generator and unsupervised model in the semi-supervised model using a GAN strengthen each other to learn the hidden statistic patterns in all the data, both labeled and unlabeled as frustration. Thus, the shared layers between the supervised and unsupervised models in the semi-supervised model using a GAN brought the advantage of learning more about the data and its statistical patterns. As a result, our approach is able to predict the frustration using only a few labeled data.

5 CONCLUSION

In this paper, a human study with fifteen users was completed. In the study, the SmartOS application collected multi-modal data from the users' computers and sent the data to AWS data servers. The collected data were used to predict human frustrations with their computers. Thus different machine learning model architectures were evaluated, and a semi-supervised model using a generative adversarial network (GAN) resulted in the highest accuracy of 90%.

REFERENCES

[1] Bidyut Bikash Hazarika, Mohammadreza Mousavizadeh, and Mike Tarn. 2019. A comparison of hedonic and utilitarian digital products based on consumer evaluation and technology frustration. *JISTEM-Journal of Information Systems and Technology Management* 16 (2019).
[2] Scott Brave and Cliff Nass. 2007. Emotion in human-computer interaction. In *The human-computer interaction handbook*. CRC Press, 103–118.
[3] Linqin Cai, Yaxin Hu, Jiangong Dong, and Sitong Zhou. 2019. Audio-textual emotion recognition based on improved neural networks. *Mathematical Problems in Engineering* 2019 (2019).
[4] George Chalhoub, Martin J Kraemer, Norbert Nthala, and Ivan Flechais. 2021. "It Did Not Give Me an Option to Decline": A Longitudinal Analysis of the User Experience of Security and Privacy in Smart Home Products. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 555, 16 pages. https://doi.org/10.1145/3411764.3445691
[5] O Chapelle, B Schölkopf, and A Zien. 2006. Semi-Supervised Learning Cambridge.
[6] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A Bharath. 2018. Generative adversarial networks: An overview. *IEEE signal processing magazine* 35, 1 (2018), 53–65.
[7] Théo Deschamps-Berger, Lori Lamel, and Laurence Devillers. 2021. End-to-end speech emotion recognition: challenges of real-life emergency call centers data recordings. In *2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–8.
[8] Clayton Epp, Michael Lippold, and Regan L Mandryk. 2011. Identifying emotional states using keystroke dynamics. In *Proceedings of the sigchi conference on human factors in computing systems*. 715–724.
[9] Paul Freihaut, Anja S Göritz, Christoph Rockstroh, and Johannes Blum. 2021. Tracking stress via the computer mouse? Promises and challenges of a potential behavioral stress marker. *Behavior Research Methods* 53, 6 (2021), 2281–2301.
[10] Surjya Ghosh. 2017. Emotion-aware computing using smartphone. In *2017 9th International Conference on Communication Systems and Networks (COMSNETS)*. IEEE, 592–593.
[11] Sepideh Goodarzy, Maziyar Nazari, Richard Han, Eric Keller, and Eric Rozner. 2021. SmartOS: Towards Automated Learning and User-Adaptive Resource

Allocation in Operating Systems. In *Proceedings of the 12th ACM SIGOPS Asia-Pacific Workshop on Systems* (Hong Kong, China) *(APSys '21)*. Association for Computing Machinery, New York, NY, USA, 48–55. https://doi.org/10.1145/3476886.3477519
[12] Hatice Gunes and Maja Pantic. 2010. Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. In *International conference on intelligent virtual agents*. Springer, 371–377.
[13] Vedika Gupta, Stuti Juyal, Gurvinder Pal Singh, Chirag Killa, and Nishant Gupta. 2020. Emotion recognition of audio/speech data using deep learning approaches. *Journal of Information and Optimization Sciences* 41, 6 (2020), 1309–1317.
[14] Ebba Håkansson and Elizabeth Bjarnason. 2020. Including human factors and ergonomics in requirements engineering for digital work environments. In *2020 IEEE First International Workshop on Requirements Engineering for Well-Being, Aging, and Health (REWBAH)*. IEEE, 57–66.
[15] Pavol Harár, Radim Burget, and Malay Kishore Dutta. 2017. Speech emotion recognition with deep learning. In *2017 4th International conference on signal processing and integrated networks (SPIN)*. IEEE, 137–140.
[16] Martin Thomas Hibbeln, Jeffrey L Jenkins, Christoph Schneider, Joseph Valacich, and Markus Weinmann. 2017. How is your user feeling? Inferring emotion through human-computer interaction devices. *Mis Quarterly* 41, 1 (2017), 1–21.
[17] Preeti Khanna and Mukundan Sasikumar. 2010. Recognising emotions from keyboard stroke pattern. *International journal of computer applications* 11, 9 (2010), 1–5.
[18] Jonathan Lazar, Adam Jones, and Ben Shneiderman. 2006. Workplace user frustration with computers: An exploratory investigation of the causes and severity. *Behaviour & Information Technology* 25, 03 (2006), 239–251.
[19] Margaret Lech, Melissa Stolar, Christopher Best, and Robert Bolia. 2020. Real-time speech emotion recognition using a pre-trained image classification network: Effects of bandwidth reduction and companding. *Frontiers in Computer Science* 2 (2020), 14.
[20] Po-Ming Lee, Wei-Hsuan Tsui, and Tzu-Chien Hsiao. 2015. The influence of emotion on keyboard typing: an experimental study using auditory stimuli. *PloS one* 10, 6 (2015), e0129056.
[21] Mao Li, Bo Yang, Joshua Levy, Andreas Stolcke, Viktor Rozgic, Spyros Matsoukas, Constantinos Papayiannis, Daniel Bone, and Chao Wang. 2021. Contrastive unsupervised learning for speech emotion recognition. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6329–6333.
[22] Zheng Lian, Jianhua Tao, Bin Liu, Jian Huang, Zhanlei Yang, and Rongjun Li. 2020. Context-Dependent Domain Adversarial Neural Network for Multimodal Emotion Recognition.. In *INTERSPEECH*. 394–398.
[23] Yu Liang, Jinheng Li, Rachata Ausavarungnirun, Riwei Pan, Liang Shi, Tei-Wei Kuo, and Chun Jason Xue. 2020. Acclaim: Adaptive Memory Reclaim to Improve User Experience in Android Systems. In *2020 USENIX Annual Technical Conference (USENIX ATC 20)*. USENIX Association, 897–910. https://www.usenix.org/conference/atc20/presentation/liang-yu
[24] Jean-Pierre Lozi, Baptiste Lepers, Justin Funston, Fabien Gaud, Vivien Quéma, and Alexandra Fedorova. 2016. The Linux Scheduler: A Decade of Wasted Cores. In *Proceedings of the Eleventh European Conference on Computer Systems* (London, United Kingdom) *(EuroSys '16)*. Association for Computing Machinery, New York, NY, USA, Article 1, 16 pages. https://doi.org/10.1145/2901318.2901326
[25] Surbhi Madan, Monika Gahalawat, Tanaya Guha, and Ramanathan Subramanian. 2021. Head Matters: Explainable Human-centered Trait Prediction from Head Motion Dynamics. In *Proceedings of the 2021 International Conference on Multimodal Interaction*. 435–443.
[26] Edmilson Morais, Ron Hoory, Weizhong Zhu, Itai Gat, Matheus Damasceno, and Hagai Aronowitz. 2022. Speech emotion recognition using self-supervised features. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6922–6926.
[27] Avar Pentel. 2015. Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State.. In *UMAP Workshops*.
[28] Mahwish Pervaiz and Tamim Ahmed Khan. 2016. Emotion recognition from speech using prosodic and linguistic features. *International Journal of Advanced Computer Science and Applications* 7, 8 (2016).
[29] Koustuv Saha, Yozen Liu, Nicholas Vincent, Farhan Asif Chowdhury, Leonardo Neves, Neil Shah, and Maarten W. Bos. 2021. AdverTiming Matters: Examining User Ad Consumption for Effective Ad Allocations on Social Media. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 581, 18 pages. https://doi.org/10.1145/3411764.3445394
[30] Sergio Salmeron-Majadas, Olga C Santos, and Jesus G Boticario. 2014. Exploring indicators from keyboard and mouse interactions to predict the user affective state. In *Educational Data Mining 2014*.
[31] Mehmet Cenk Sezgin, Bilge Günsel, and Güneş Karabulut Kurt. 2011. A novel perceptual feature set for audio emotion recognition. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 780–785.
[32] Mandeep Singh and Yuan Fang. 2020. Emotion recognition in audio and video using deep neural networks. *arXiv preprint arXiv:2006.08129* (2020).
[33] Oren Wright. 2019. *Emotion Recognition from Voice in the Wild*. Technical Report. CARNEGIE-MELLON UNIV PITTSBURGH PA PITTSBURGH United States.
[34] Takashi Yamauchi, Anton Leontyev, and Moein Razavi. 2019. Assessing emotion by mouse-cursor tracking: Theoretical and empirical rationales. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 89–95.
[35] Takashi Yamauchi and Kunchen Xiao. 2018. Reading emotion from mouse cursor motions: Affective computing approach. *Cognitive science* 42, 3 (2018), 771–819.
[36] Philippe Zimmermann, Sissel Guttormsen, Brigitta Danuser, and Patrick Gomez. 2003. Affective computing—a rationale for measuring mood with mouse and keyboard. *International journal of occupational safety and ergonomics* 9, 4 (2003), 539–551.