Detection and analysis of a programmer’s confidence
A multimodal approach

Alexandru Ionescu
Computer Science
Babeş-Bolyai University
Cluj-Napoca, România
andrei.alexandru.ionescu@stud.ubbcluj.ro

Gabriel Mărginean
Computer Science
Babeş-Bolyai University
Cluj-Napoca, România
laurentiu.marginean@stud.ubbcluj.ro

Miruna Hancu
Computer Science
Babeş-Bolyai University
Cluj-Napoca, România
miruna.hancu@stud.ubbcluj.ro

Fivia Lele
Computer Science
Babeş-Bolyai University
Cluj-Napoca, România
fivia.lele@stud.ubbcluj.ro

ABSTRACT:

A key element that stays behind the great technology evolution we are witnessing in today’s world are the programmers. We can say that developing programmers’ skills contributes to pushing the wheel further on. Therefore, we have conducted an experiment to better understand the deficiencies in programmers’ coding capabilities that can be used in order to help them fill the gaps. The research is based on detecting a programmer's confidence. Its applicabilities are diverse including the academic scenery, extending up to technical interview use. We propose a non-intrusive approach of assessing this metric based on keyboard and mouse input data, as well as compilation feedback (errors, warnings).

CCS CONCEPTS:

- Human-computer interaction (HCI)
- Artificial intelligence
- Life and medical sciences

KEYWORDS:

Keystroke analysis, Confidence identification, Log compiler errors.

1 INTRODUCTION

The significance of emotion in maintaining mental health has led to increased interest in research that revolves around this subject. The impact of these studies has spread onto many scientific branches, computer science included.

To tackle this issue, it is particularly important to note that several methods for identifying user emotions have been probed before: [2] voice intonation analysis, facial expression analysis, physiological sensors attached to the skin, and thermal imaging of the face. Even if these surveys have been successful, they still display one or both of the two main problems preventing wide-reaching use: they are intrusive to the user, or they demand specialized apparatus that is expensive and not found in a typical home or office environment.

Keystroke analysis and mouse movements are unique biometrical features, which can unveil the emotions of the user, such as happiness, anger, content, or confidence. Their advantage encompasses a friendly, long-term use, and non-intrusive approach. Our study is aiming to detect the level of confidence of a user, by using a thorough analysis of keystroke, mouse movements, and program execution data.

Building on the idea that [3] “nowadays, graphics and computing capabilities are much more powerful and stronger,” the question must be addressed: Is it possible, as software developers, to build a model strong-enough so as to identify coding confidence real-time?

It is known that a method exploiting normal daily human-computer interactions is attractive, especially in times like nowadays when human relations have been altered by the pandemic situation.

If we are considering the educational context, the connections between students or between students and teachers are jeopardized by Covid-19's restrictions. The online environment creates an artificial social context, leaving the tutors in a situation where they’re unable to correctly identify the students’ level of confidence, which is usually exposed in physical interaction between them, through different aspects such as facial features, micro-expression, voice inflections or any other form of non-verbal communication.

Thus, our application would surely be of use in this domain; the teachers being able to use it as a tool for collecting data and after that giving personalized feedback. Also, the data generated can be of use when a tutor wants to verify how efficient his/her lecture was.

Measuring the confidence of a programmer while writing code could prove useful in a live coding interview setting, where the interviewer can take into account, among other skills and domain knowledge, this metric as well. Oftentimes, multiple candidates meet the required criteria, and might be difficult to choose between them and this additional information available might differentiate between candidates and help interviewers to better distinguish between them.

The approach we tend towards is one in which, while coding, the user’s “movements” on the keyboard and handling the mouse are being “recorded” in a non-intrusive manner. The data will be further interpreted and transformed so as to show an intelligible result, relevant and easily formulated, meaning the percentage of identified confidence.

In the previous literature, some repetitive characteristics are being considered for identifying and analyzing emotion: keystroke duration, keystroke latency, and keyboard typing, frequency of moving the mouse. To facilitate benchmarking which can reflect near real-world scenarios, we want to stick to these manners of research, but also broaden our horizons by introducing, as an element of novelty, the study of a programmer’s confidence by counting the number of compilation errors found in each session of recorded coding time. With the support that we find in the previous work, we can rely on saying that a complex union of the different stimuli affects the state of the subjects we study. Thus, we want to record the moments when he/she feels confident and the factors that may lead to this.

2 RELATED WORK

The analysis of peripheral hardware as a way of identifying the user and his/her emotions has been an active area of research recently. Many studies, technologies, and tools have been built on this premise, and we retrieve this in the following form: user authentication based on keystroke dynamics, recognition of [1] emotional states, [2] confidence, [5,10] stress, based on mouse movements/ keystroke dynamics, [11] mobile keystroke.

Our study is based on the analysis of the keystroke dynamics and compiler errors of users, to identify increases and decreases in their level of confidence. Peripheral input is an automatic, non-intrusive approach with a reduced cost of application.

We need to be aware that building an emotion recognition system is a challenging task, in the means of how we gather and represent data, or train our systems. In the following paragraphs, we will describe the approaches suggested for use in previous studies.

Ref.	Emotion analyzed	Peripheral	Analysis Text Type	Method	Results
[2]	Confidence (+others)	keystroke	Free+Fixed Text	Decision Tree	Classification accuracy: 86.2% for fixed text, Free text results ignored
[3]	Positive and negative	keystroke + facial feedback	Fixed Text	Statistical analysis	Significant differences in the typing patterns under positive and negative emotions for all subjects.
[4]	Looking for any influence on mood and performance	keystroke	Fixed text	Two-way Valence (3) x Arousal (3) ANOVAs	The effect of arousal is significant in keystroke duration ($p < .05$), keystroke latency ($p < .01$), but not in the accuracy rate of keyboard typing.
[5]	Stress	keystroke + language parameters	Free text	Decision trees, support vector machine, k-NN, AdaBoost, neural networks	75.0% for cognitive stress (k-NN), 62.5% for physical stress (AdaBoost,

					SVM, neural networks) conclusion: number of mistakes made while typing decreases under stress
[6]	Mapping personality traits	keystroke	Fixed text	Neural networks	it was not possible to identify an individual's personality traits from the typing rhythm using the approaches described in this work
[10]	Stress	keystroke + mouse features	Free text	Linear Discriminant Analysis, feature vector, 5-NN, SVM, NB	Overall accuracy: 74.5%

By analyzing the previous work, we observed that the techniques which were mostly used to represent data were: decision trees, neural networks, and AdaBoost, followed by Linear Discriminant Analysis and valence arousal model.

The valence arousal model, [4, 13], supposes that emotions can be defined by a coincidence of values on two dimensions that are (AV space): arousal (intensity of the feeling, the excitement of the person.) and valence (the feeling is positive or negative.). We can find the emotion we are looking for, "confidence", as a part of this representation, as seen below in Fig.1:

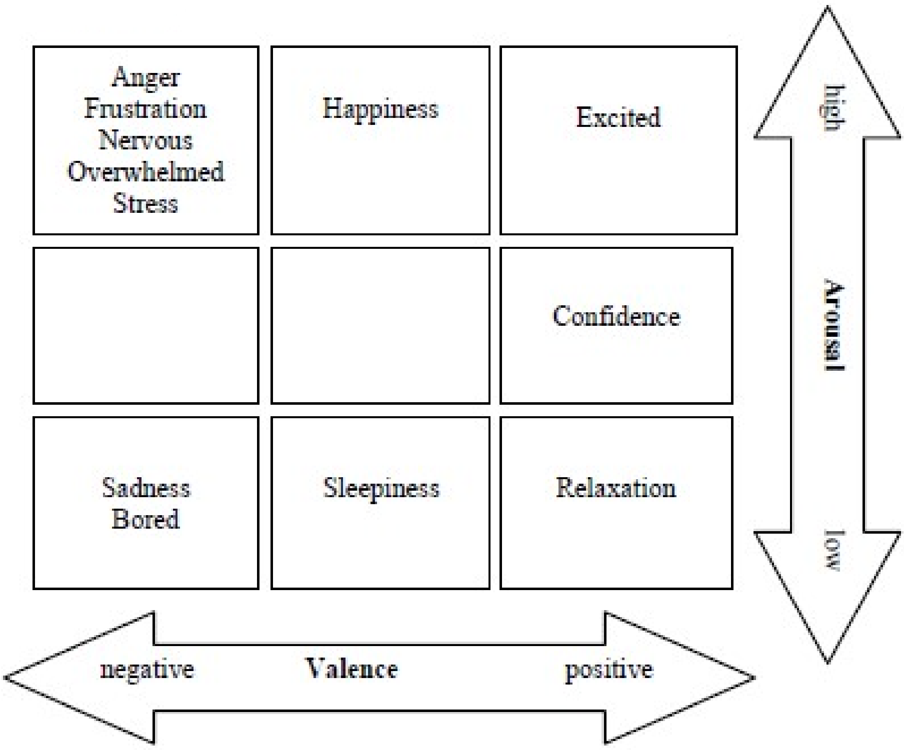

Fig.1

Knowing that the results of these studies were satisfactory, we can base our study on the solid ground they built. So, we can create a model which will identify "confidence" as a stand-alone emotion, and associate positive emotions with a high level of confidence and negative emotions with a low level of confidence.

With this, we can analyze the data previously gathered, structured in Fig.2.

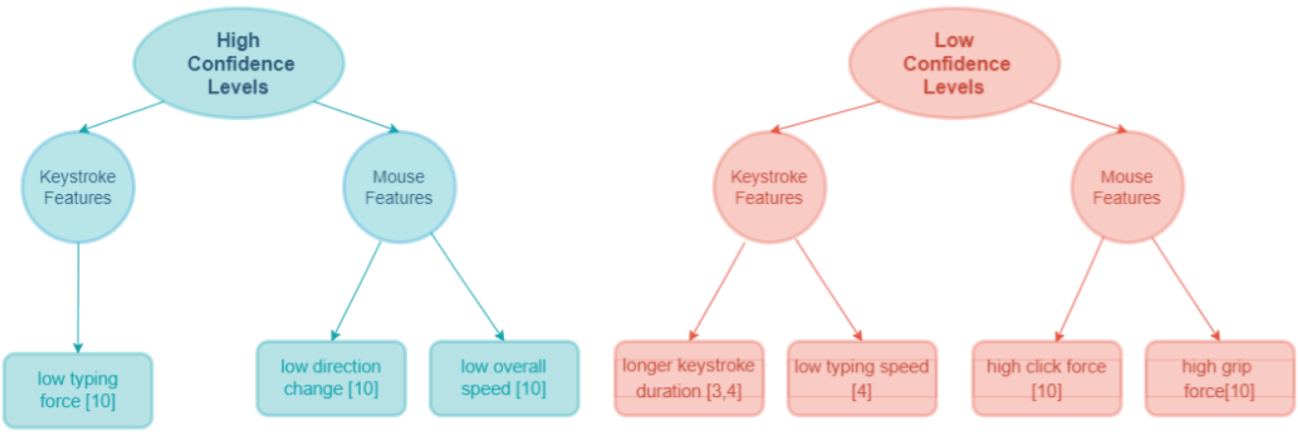

Fig.2

It was observed that the more successful keystroke studies were the ones focusing on fixed-text analysis, meaning that, at a certain moment, they asked the user to introduce a specific sequence of characters, one with which the model was already trained.

However, few studies, [2,10], which rely on free-text analysis were found. Free text analysis is a method that implies monitoring the typing of the user for a certain period, with its knowledge. It is a more difficult task to implement because an individual's typing patterns are not stable; there are factors like stress, emotional stimuli, or changes in their cognitive function that may alter their behavior, and data cannot be collected properly. We've noticed this in [2]'s results, which were not satisfactory.

Even so, with a thoroughly developed model, proper technology, and a systematic data analysis method, the authors of articles [10] managed to obtain satisfactory results on free text analysis. One of the factors that helped with this was the use of a multi-modal approach, which contained keystroke, and mouse features analysis. We observed the same phenomena in articles [5,7], with impressive results, suggesting that multimodal methods of analyzing emotions are more accurate.

Since we have all these study examples on recognizing emotion through peripheral devices implemented and discussed before, we can observe their advantages and disadvantages. We choose to take the research further and tackle confidence identification with a multimodal approach, one which implies keystroke analysis on free text and compiler errors detection.

3 EXPLORATORY EXPERIMENT

Previous studies [2,4,13] have highlighted the possibility of using keyboard typing data as well as mouse features to detect emotions. Specifically, keystroke duration, keystroke latency, the accuracy rate of the keyboard, frequency of special keys, and smoothness of the pointer paths.

We designed an experiment to gather the aforementioned data while subjecting the participants to different circumstances.

The study consisted of two sessions, each session performed on the same day. Fig. 3 shows the structure of each session.

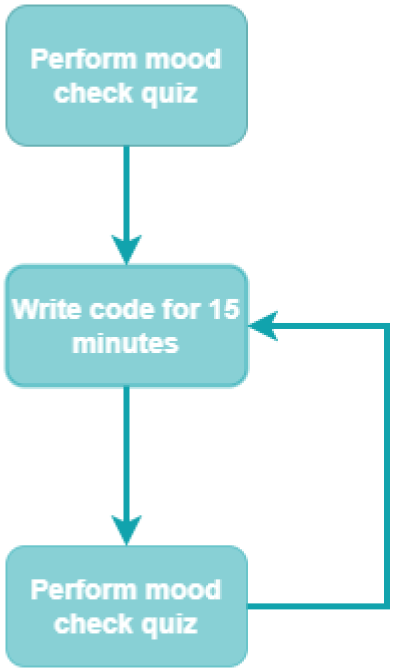

Fig. 3

First, the participants were asked to fill out a questionnaire about their arousal and valence at that moment. We also asked them to provide information about their programming background (how long they have been programming and what studies they have in the domain).

The study is divided into two blocks: the control and experimental block. Before starting the study, the participants performed a warm up session for 5 minutes consisting of trivial tasks, followed by two code writing sessions.

The participants were assigned problems based on their skills and knowledge.

During the control block, our goal was to induce a high level of confidence by providing the users simple and trivial tasks to solve, while in the experimental block they performed more challenging and thoughtful tasks in order to reduce their confidence. Once participants finished the first block, they were allowed a 15-minute break before proceeding to the next block.

Once every 15 minutes the participants were required to report their perceived valence, arousal, and confidence level by filling out a questionnaire which served two purposes.
First, it allowed us to determine whether or not the levels of confidence were successfully induced, and second, it provides an opportunity to analyze changes in mouse and keystroke behavior elicited by the prior priming task by comparing the data collected during an easy (control) session with the data collected during the difficult (experimental) session.

For self-reported valence, arousal, and confidence we used the 5-Point Self-Assessment Manikin (Fig. 4), which has been extensively used for self-reporting arousal and valence, along with two additional questions:
"How mentally demanding was the task?" and *"How confident are you that the task was solved correctly?"* with a 5-Point precision.

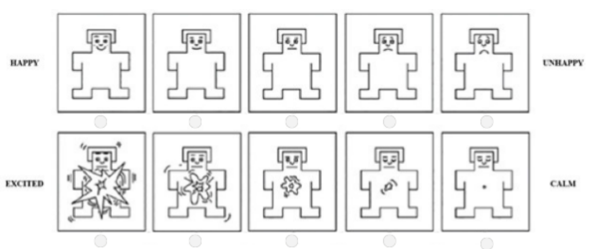

Fig. 4

4 RESULTS AND STUDY

The study consisted of 24 participants, out of which 2 had to be disqualified due to improper data collection. The persons were eligible for successfully completing both of the two blocks of the experiment, meaning they were disposing of the necessary tools (PC, keyboard, mouse, and the command-line utility) and they were possessing the basic computer science knowledge in order to fulfill the coding challenges. Most participants were Computer Science students or professionals from the IT industry with at least 2-5 years of experience because our research is targeting the programming world and we wanted the results to be as accurate and relevant in finding a solution as possible. They were all relatively young adults, with ages varying between 20 and 40.

The subjects had to go through the experiment's phases, the quiz, and then the solving of the problems. The metrics recorded were the number of keystrokes, the average keystroke duration, the number of mouse clicks, the number of backspaces or delete key presses, and the number of program executions, all of which have been extracted from sessions of 15 minutes of coding.

After combining the recorded metrics and quiz results with the help of a decision tree, the result is computed as a number between 1 and 5, with 1 indicating a lack of confidence, and 5 indicating high confidence. To compute the model's accuracy, these predictions were then compared against the participants' self-reported confidence during the quiz.

From analyzing the results we observed a naturally strong correlation between high levels of confidence detected by our system and the following 2 factors: a calm state and feeling that the task was not mentally demanding. Furthermore, we could also deduct a correlation between the decreased number of backspace key-press events and high levels of confidence.

We have gathered data from approximately 30 hours of recording and the accuracy of the solution is 68.72%. The input data from the database can be largely extended, thus conducting to greater levels of accuracy on behalf of the algorithm's behavior.

5 CONCLUSION

Our study researched if by keyboard features, mouse clicks, and program execution data we can correctly identify the moments when our user is feeling more or less confident.

We built a multimodal, non-intrusive approach, that was adapted into an experiment in which the users performed a simple programming task and a more complicated one, which required the use of a keyboard and a compiler. The analysis was made on live written code, meaning a free text analysis, thus addressing the limitations found in the literature [6,10].

The above-mentioned experiment was run on 24 candidates, out of which 2 were not eligible, due to improper data collection.

With identifying subtle changes in keyboard typing style during the assigned tasks, namely the number of pushed down keystrokes, the pressure time, and latency, the number of backspace presses or deletes, accumulated with the number of mouse clicks and the number of runs of the compiler; we were able to obtain an overall accuracy of: 68.4% in identifying confidence, using decision trees.

In comparison with [10], we improved the research by creating a more user-friendly environment, since the experiment was not realized in a laboratory setting. Our model can be used from home, school, or work, without any restrictions, and is a useful tool for one's confidence level identification, thus a mood improver.

By using a multimodal approach, we obtained a result similar to the one in the article [10] on free text analysis.

However, we encountered some limitations, caused by the small dataset that we ran our experiment on. Finally, we observed that if we enlarge the dataset with more values, the accuracy increases significantly, letting us see hope for future improvements and studies.

Multimodal approaches tend to be more accurate than single-oriented ones, and we built our idea on this premise, with a solid base on the previous work mentioned. We had satisfactory results, helping us to get a step closer to more non-intrusive, accessible methods of confidence detection.

REFERENCES:

[1] A. Kołakowska, *"A review of emotion recognition methods based on keystroke dynamics and mouse movements",* 2013 6th International Conference on Human System Interactions (HSI), 2013, pp. 548-555, doi: 10.1109/HSI.2013.6577879.

[2] Epp, Clayton, and Lippold, Michael and Mandryk, Regan L, *"Identifying Emotional States Using Keystroke Dynamics"*, 2011, 9781450302289, Association for Computing

Machinery, New York, NY, USA, https://doi.org/10.1145/1978942.1979046.

[3] W. Tsui, P. Lee and T. Hsiao, *"The effect of emotion on keystroke: An experimental study using facial feedback hypothesis",* 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 2870-2873, doi: 10.1109/EMBC.2013.6610139.

[4] Lee PM, Tsui WH, Hsiao TC, *"The Influence of Emotion on Keyboard Typing: An Experimental Study Using Auditory Stimuli",* 2015 PLOS ONE 10(6): e0129056. https://doi.org/10.1371/journal.pone.0129056

[5] Lisa M. Vizer, Lina Zhou, Andrew Sears, "*Automated stress detection using keystroke and linguistic features: An exploratory study, International Journal of Human-Computer Studies",* 2009, Volume 67, Issue 10, Pages 870-886, ISSN 1071-5819, https://doi.org/10.1016/j.ijhcs.2009.07.005.

[6] Goulart, F. and Dantas, D. (2021). "*Mapping Personality Traits through Keystroke Analysis"*, 2021, In Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 2: ICEIS, ISBN 978-989-758-509-8; ISSN 2184-4992, pages 474-482. DOI: 10.5220/0010456304740482

[7] L. Yang and S. -F. Qin, *"A Review of Emotion Recognition Methods From Keystroke, Mouse, and Touchscreen Dynamics*", in IEEE Access, vol. 9, pp. 162197-162213, 2021, doi: 10.1109/ACCESS.2021.3132233.

[8] Freihaut, P. and Göritz, A.S., 2021. "*Does Peoples' Keyboard Typing Reflect Their Stress Level?."* Zeitschrift für Psychologie.

[9] Yaacob, Mohd Noorulfakhri & Syed Idrus, Syed Zulkarnain & Mustafa, Wan & Jamlos, Mohd & Abd Wahab, Mohd Helmy. (2021). "*Identification of the Exclusivity of Individual's Typing Style Using Soft Biometric Elements."* Annals of Emerging Technologies in Computing. 5. 10-26. 10.33166/AETiC.2021.05.002.

[10] D. R. Dacunhasilva, Z. Wang and R. Gutierrez-Osuna, *"Towards Participant-Independent Stress Detection Using Instrumented Peripherals"*, 2021, in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2021.3061417.

[11] Maiorana, E., Kalita, H. and Campisi, P., 2021. "*Mobile keystroke dynamics for biometric recognition: An overview"*. IET Biometrics.

[12] Tsimperidis, I., Yucel, C. & Katos, V., 2021. "*Age and Gender as Cyber Attribution Features in Keystroke Dynamic-Based User Classification Processes*". Electronics, 10(7), p.835. Available at: http://dx.doi.org/10.3390/electronics10070835

[13]Solanki, ,R. and Shukla, P. (2014). "*Estimation of the user's emotional state by keystroke dynamics"*. International Journal of Computer Applications, 94(13).