Modelling Experience of Users during IDE Usage via Behavioural Biometrics

Daniela Chudá, Lukáš Hagara, Kamil Burda

Abstract: *In an e-learning environment involving programming courses, it is difficult for teachers to provide help to their students due to restricted communication. In this paper we propose to automatically evaluate the level of experience of users in integrated development environments (IDEs) using behavioural biometrics extracted from the mouse and keyboard. Based on our experiment, we defined three experience levels and subsequently evaluated the accuracy of our proposed method using two groups of mouse biometrics and keyboard biometrics. The results show that keystroke dynamics have a great potential of accurately assessing the level of experience of users during IDE usage.*

Key words: *integrated development environment, behavioural biometrics, user model, experience*

INTRODUCTION

In the education of programming, it is indispensable for teachers to evaluate the level of experience of users of integrated development environments (hereinafter IDEs). Any knowledge about the users can aid teachers in better assessing their needs and provide more appropriate assistance to them. In an e-learning environment, such as online courses, these efforts are hindered by the restricted teacher-student communication.

The level of experience of users[20] can be potentially evaluated automatically based on the user activities with input devices, namely mouse and keyboard. In other words, we can measure the experience level of users based on behavioural biometrics acquired from input devices when performing tasks in IDEs.

The goal of this paper is to determine the feasibility of modelling the experience of users based on the behavioural biometrics extracted from keyboard and mouse usage in IDEs. Section **RELATED WORK** discusses user modelling based on behavioural biometrics and the usage of biometrics in tasks such as user identification and authentication. Section **METHOD** describes our proposed method to evaluate the level of experience of users. Section **EVALUATION OF EXPERIENCE LEVEL OF USERS** describes the experiment conducted and the method to pre-determine the experience level of users based on a survey. The results are presented in the section **RESULTS**. Section **CONCLUSIONS AND FUTURE WORK** concludes the paper and discusses future endeavours.

RELATED WORK

Behavioural biometrics allow us to observe users' behaviour based on their usage patterns of input devices, such as keyboard (keystroke dynamics) or mouse (mouse movements, mouse clicks). Behavioural biometrics can be used for a variety of purposes, including user identification and authentication or recognizing emotions. Given their relatively insufficient accuracy, behavioural biometrics are used as an additional factor to the traditional login-and-password user identification and authentication, e.g. in the form of continuous user authentication. While physiological biometrics (fingerprint, iris, DNA) achieve better accuracy, they are not widely used due to special hardware requirements [9].

Performing any task (user identification and authentication, and in our case modelling experience of users) using biometrics involves modelling the user. First, raw data are gathered from input devices, such as keystroke events from keyboards (key up, key down) and mouse clicks and movements from mice, with each action being associated with a time stamp. The raw data are subsequently pre-processed (performing normalization, splitting data into samples) and biometrics are extracted from the data, e.g. dwell time for

---

20 We use the term "experience of users" instead of "user experience" to avoid confusion with the concept of user interface usability.

key strokes or average mouse velocity. The biometrics represent the template of a user, which is subsequently used in finding the best matching template for newly extracted biometrics. Popular template-matching methods include the Support Vector Machine (SVM), decision trees, or *k*-Nearest Neighbours. False Acceptance Rate (FAR), False Rejection Rate (FRR) and Equal Error Rate (EER) are frequently used to evaluate the accuracy of a biometric system.

Mouse and keyboard biometrics have been used more frequently in scenarios requiring pre-defined text input (such as username and password during authentication) [3] or movements [7][1], also referred to as static input. Typing paragraphs of text or performing unconstrained mouse movements generate dynamic input, which are more difficult to evaluate [3][6][8]. For example, free text input requires comparing a sequence of *n* characters (also referred to as *n*-graphs) and longer period of data gathering due to the variation of the text contents. Compared to static input, dynamic input better represents natural user behaviour [3], such as in the workplace [5].

In [4], authors examined keystroke dynamics during password typing. Flight time and dwell time were extracted between each pair of consecutive keys. With only 4-5 samples and one-class SVM, EER of 11.93% was achieved. A mouse-based standalone authentication method is proposed in [1], where users were prompted to perform mouse movements in a pre-defined pattern. Using the mean, standard deviation, minimum and maximum of velocity, deviation from straight line, angle and positive acceleration, authors were able to achieve EER of 5.9%. Mouse dynamics was also examined in [8] in a controlled online environment. Using angle-based biometrics and dual class SVM, EER of 1.3% was achieved. Another approach was taken in [7], where mouse movements were split to curves by two consecutive clicks. From each curve, the curve size, length, time to draw the curve, movement angle, average velocity, acceleration, click duration were extracted as biometrics. Using *k*-NN, the authors achieved 92% accuracy as the best result. Authors in [2] performed mouse-based user identification on the web using a memory game, with mouse click-based biometrics (click duration, time between click and last movement, time between click and next movement) outperforming movement-based biometrics in terms of accuracy. 50 users were successfully identified using 100 clicks with the accuracy of 85%.

METHOD

In this section we describe our proposed method to model the experience of users based on behavioural biometrics extracted from keyboard and mouse during the usage of an IDE.

For the purposes of logging raw data, we developed our own logger in Python language using the *pyHook* library capable of intercepting all key events (key pressed, key up/down, time stamp) and mouse events (left/middle/right mouse button up/down, mouse cursor *xy* coordinates, scrolling, time stamp).

The biometrics (biometric features) that we extract from the raw data form vectors of values of each biometric feature for one user. We normalize the values of each feature to interval $< -1, 1 >$. In order for each feature to map to the entire normalized interval, we first determine the minimum and maximum value of each feature separately.

In our proposed method, we train and evaluate models of experience of users using three different groups of biometrics: curve-based biometrics from mouse movements (based on [7]), angle-based biometrics from mouse movements (based on [8]) and keystroke-based biometrics from keyboard. The biometrics are shown in Tables 2, 3 and 4, respectively.

For the curve-based biometrics, we split the data into curves by a single mouse click. If the curve was ended by a different event than a mouse click or the duration of the curve exceeded a given threshold, we discarded the curve. Next, we grouped six consecutive curves into one group. From each group we extracted the mean and standard deviation of the biometrics specified in Table 2. The template width refers to the length of the feature vector used in classification. Each separate biometric has a template width of 3 – its mean, standard deviation and additionally the direction of the mouse movement. We only consider four directions in our proposed method according to quadrants in the Cartesian coordinate system.

For the angle-based mouse biometrics, we also consider other mouse movements that were discarded for the curve-based mouse biometrics, including movements ended by events other than a mouse click. The direction represents the direction of the end-to-end line from the endpoints of a curve. The angle of curvature represents the angle of three points on a curve – two endpoints and one midpoint. The curvature distance is the largest deviation of a curve from the end-to-end distance of its endpoints. The values for the direction and the angle of curvature were split into 5- and 10-bin histograms, with 250 angle values in each bin. For the curvature distance, we consider its mean and standard deviation.

For the keystroke-based biometrics, we consider the dwell time (duration of a key press), and four variants of flight time: time between two key presses, time between a key release and the press of the next key, time between two key releases, and time between a key press and the release of the next key. If the delay between two key events is longer than a given threshold, we discard such keystroke data. To extract the proposed biometrics, we consider 40 successive key strokes (a half of the conventional maximum 80 characters per a source code line). It is not feasible to extract the biometrics for each key separately due to memory constraints and the sparsity of the values in 40 key strokes, therefore we categorize the keys to 8 groups as shown in Figure 8. The number in each key represents the scan code of that key. All keys with scan code greater than 88 or equal to 0 form one group. The length of the feature vector for flight-time features is $8^2 = 64$ due to tracking all combinations of two consecutive key presses. For the dwell time, we consider each key separately, hence we get a feature vector of length 90 (the number of possible scan codes).

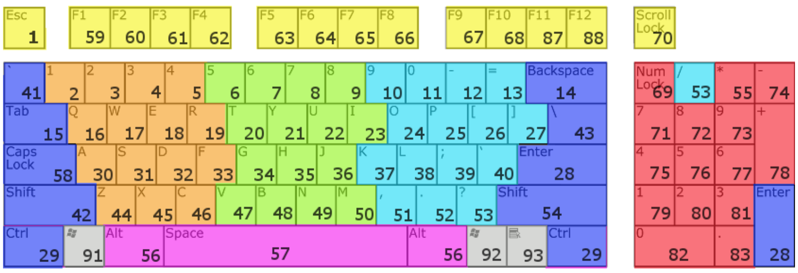

Figure 8: Grouping keys for keystroke-based biometrics

The extracted and normalized biometric features form a data set to model the experience of users. We evaluate the experience by considering the data from all users, i.e. we do not distinguish the experience of individual users. For template matching (classification), we split the data set into training and test sets in the ratio of 2:1. For classification, we chose the Support Vector Machine (SVM) to identify the correct experience level (class) as it generally provides good classification accuracy and is frequently used in other research works and experiments. While SVM is a binary classifier and we need to distinguish three classes, we use the *LIBSVM* library to perform classification, which is capable of performing multi-class classification by binarizing the classification task. For more accurate results, the number of positive and negative samples should be in ratio of 1:1, meaning we remove samples to achieve the desired ratio in a random manner. Because the classification will perform on different samples on each run, we consider the average classification results from multiple runs. We measure the accuracy of our method by the classification accuracy, the False Acceptance Rate (FAR, equivalent to false positive rate) and the False Rejection Rate (FRR, equivalent to false negative rate).

EVALUATION OF EXPERIENCE LEVEL OF USERS

We performed a controlled experiment with 23 students of informatics, of which 20 were bachelor degree students and 3 were master or doctoral degree students. The participants were instructed to perform tasks in Microsoft Visual Studio IDE using keyboard and mouse, such as writing procedures or splitting a single source file into multiple source files. The length of the experiment was 38 minutes in average, in the range of 17-55 minutes.

After finishing the tasks, each user was asked to fill out a survey that included 12 questions about IDE features and usage of the IDE by the user and information such as age, years of study in informatics and self-assessment of experience of working in the IDE.

The survey was used to estimate the experience level of each user and define an appropriate number of experience levels to which the users would be classified. During the survey evaluation, we removed 4 questions in which a vast majority of users provided the same answer. Each question was assigned a score, with four-choice questions being given 0-3 points in the ascending order for each answer. The overall experience score of each user is computed as a sum of points gained from each question.

Based on the experience score, we defined three experience levels of users: inexperienced, moderately experienced and experienced. Each experience level is designed such that each level covers approximately 1/3 of experience scores from all users. Using the cumulative distribution function of experience scores, we determined the following intervals of experience score for each experience level: $(-\infty, 8 >$ for inexperienced, $(8, 11 >$ for moderately experienced and $(11, \infty)$ for experienced.

RESULTS

In this section we describe the results of the experiment to model the experience level of users using curve-based mouse biometrics, angle-based mouse biometrics and keyboard biometrics. For mouse biometrics, the raw data were gathered every 15 milliseconds.

Table 2 shows the classification accuracy (ACC), FAR and FRR of experience modelling for each biometric and several combinations thereof for the curve-based mouse biometrics. When splitting logged data to curves, threshold of 0.5 seconds was used as the maximum allowed duration of a curve before a mouse click event. The click duration as a standalone feature and the combination of acceleration, time on curve, click duration and velocity achieved the best classification accuracy.

Table 2: Results for experience levels for curve-based mouse biometrics

Feature	Class Prediction [%]			Template width
	ACC	FAR	FRR	
Acceleration (A)	23.320	42.029	34.651	3
Number of points constituting a curve (N)	21.269	45.274	33.458	3
Time to complete a curve (T)	24.498	41.098	34.404	3
Curve length (CL)	19.018	44.584	36.398	3
Click duration (CD)	42.474	34.426	23.100	3
Velocity (V)	28.825	39.891	31.284	3
T + CL + CD	43.769	33.131	23.100	7
T + CD	43.884	32.722	23.394	5
A + T + CD + V	44.109	33.535	22.356	9

Table 3 shows the results for angle-based mouse biometrics. The direction achieved dissatisfactory results, while the angle of curvature with 10 bins and the curvature distance showed promising results. While the curvature distance performed slightly worse, the significantly shorter template width can be beneficial in classification performance optimization. Combining the 10-bin angle of curvature and the curvature distance considerably increased the classification accuracy.

Table 3: Results for experience levels for angle-based mouse biometrics

Feature	Class Prediction [%]			Template width
	ACC	FAR	FRR	
Direction (5)	23.6016	39.9727	36.4256	5
Direction (10)	23.8606	41.0188	35.1206	10
Angle of curvature (5)	23.6111	38.8889	37.5	5
Angle of curvature (10)	38.5093	31.677	29.8137	10
Curvature distance	37.7451	35.2941	26.9608	2
D (10) + A (10) + C	45.5446	27.3927	27.0627	22
A (10) + C	47.351	27.1523	25.4967	12

Table 4 shows the classification results for keystroke-based biometrics. The maximum allowed delay between two key events was set to 1 second. Dwell time and combinations of dwell time with select flight-time features provide satisfactory classification accuracy above 65%. Thus, keystroke-based biometrics significantly outperform mouse biometrics in terms of accuracy, FAR and FRR.

Table 4: Results for experience levels for keystroke-based biometrics

Feature	Class Prediction [%]			Template width
	ACC	FAR	FRR	
Dwell Time	67.518	19.708	12.774	90
Flight Time 1	34.181	37.853	27.966	64
Flight Time 2	39.193	36.599	24.207	64
Flight Time 3	33.795	39.058	27.147	64
Flight Time 4	34.000	37.143	28.857	64
DT + F4	65.756	19.854	14.390	154
DT + F2 + F4	66.486	20.290	13.225	218
DT + F3	68.000	20.000	12.000	154

CONCLUSIONS AND FUTURE WORK

In this paper we proposed an automatic method to aid in evaluation of experience of users in integrated development environments using behavioural biometrics extracted from keyboard and mouse. Our method is based on dynamic input, i.e. we observe the natural behaviour of the users instead of constraining their mouse movements or key strokes.

The results show that it is feasible to use our method to help better evaluate the level of experience of users, especially with keystroke dynamics that achieved 68% accuracy in recognizing the correct experience level of users. The results can be used to improve the teacher-student relationship by allowing the teacher to provide more accurate help, especially in online courses with the direct lack of social contact between the teacher and the student. The results can also be used to let IDEs automatically provide help to the user or adapt the user interface for improved usability.

In the future, we aim to revise the evaluation of mouse biometrics to improve their accuracy and further improve the accuracy of the dwell time feature in keystroke dynamics (which as a standalone feature proved to be relatively very accurate in distinguishing experience level). We also plan to conduct additional experiments and gather more data sets for more accurate evaluation of levels of experience and possibly defining additional experience levels.

REFERENCES

[1] Y. Aksari and H. Artuner, "Active authentication by mouse movements," in *24th International Symposium on Computer and Information Sciences, 2009. ISCIS 2009*, 2009, pp. 571–574.

[2] D. Chudá, P. Krátky, and J. Tvarožek, "Mouse Clicks Can Recognize Web Page Visitors!," in *Proceedings of the 24th International Conference on World Wide Web*, Republic and Canton of Geneva, Switzerland, 2015, pp. 21–22.

[3] P. Kang and S. Cho, "Keystroke Dynamics-based User Authentication Using Long and Free Text Strings from Various Input Devices," *Inf Sci*, vol. 308, no. C, pp. 72–93, Jul. 2015.

[4] Y. Li, B. Zhang, Y. Cao, S. Zhao, Y. Gao, and J. Liu, "Study on the BeiHang Keystroke Dynamics Database," in *2011 International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–5.

[5] F. Monrose and A. D. Rubin, "Keystroke Dynamics As a Biometric for Authentication," *Future Gener. Comput. Syst.*, vol. 16, no. 4, pp. 351–359, Feb. 2000.

[6] P. S. Teh, A. B. J. Teoh, and S. Yue, "A Survey of Keystroke Dynamics Biometrics," *Sci. World J.*, vol. 2013, no. 4, p. 408280, Nov. 2013.

[7] A. Weiss, A. Ramapanicker, P. Shah, S. Noble, and L. Immohr, "Mouse movements biometric identification: A feasibility study," *ResearchGate*, Jan. 2007.

[8] N. Zheng, A. Paloski, and H. Wang, "An Efficient User Verification System via Mouse Movements," in *Proceedings of the 18th ACM Conference on Computer and Communications Security*, New York, NY, USA, 2011, pp. 139–150.

[9] P. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez, "Affective Computing—A Rationale for Measuring Mood With Mouse and Keyboard," *Int. J. Occup. Saf. Ergon. JOSE*, vol. 9, no. 4, pp. 539–51, Feb. 2003.

ABOUT THE AUTHORS

Assoc. Prof. Mgr. Daniela Chudá, PhD., Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technology, Slovak University of Technology in Bratislava, Slovak Republic, Phone: +421 2 210 22 318, E-mail: daniela.chuda@stuba.sk.

Bc. Lukáš Hagara, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technology, Slovak University of Technology in Bratislava, Slovak Republic, Phone: +421 2 210 22 331, E-mail: xhagaral@stuba.sk.

Ing. Kamil Burda, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technology, Slovak University of Technology in Bratislava, Slovak Republic, Phone: +421 2 210 22 331, E-mail: kamil.burda@stuba.sk.

The paper has been reviewed.