Identifying Soft Biometric Features from a Combination of Keystroke and Mouse Dynamics

Sally Earl, James Campbell, and Oliver Buckley

School of Computing Science, University of East Anglia, Norwich, NR4 7TJ, UK
{s.earl, j.campbell1, o.buckley}@uea.ac.uk

Abstract. In this preliminary paper, we investigate the use of keystroke and mouse dynamics as a means of identifying soft biometric features. We present evidence that combining features from both provides a more accurate means of identifying all of the soft biometric traits investigated regardless of the machine learning method used. The data presented in this paper gives a thorough breakdown of accuracy scores from multiple machine learning methods and numbers of features used.

Keywords: Soft Biometrics · Keystroke Dynamics · Mouse Dynamics.

1 Introduction

While biometrics are becoming more prevalent as alternatives to conventional password and passcode systems, a key source of data which is often overlooked is the mouse. When we take into account the sheer number of times the average user moves or clicks a mouse per day, there is potentially a vast amount of data that goes unused. Utilising this data, could prove to uplift prediction scores of soft biometric features when combined with keystroke dynamics, and this is discussed in the research presented in this paper.

Previous work completed, such as that of [2] has found that mouse dynamics directly complement keystroke dynamics, resulting in a higher accuracy level when data is analysed. Combining mouse with keystroke dynamics therefore, could be a way to leverage certain parts of the users interaction with the mouse, in a way that directly assists in the authentication and identification of certain soft biometrics of a user.

In this study we aim to utilise keystroke and mouse dynamics as a method of predicting soft biometric features. Furthermore, we aim to understand if mouse dynamics can give an uplift to the accuracy of prediction in the hopes of combining these two widely used technologies. This then allowing us to hopefully establish a new use for potentially under-utilised data.

2 Literature Review

Biometrics can be broadly split into two categories, physical and behavioural. Physical biometrics are physical features of a person which can be analysed to

uniquely identify them, whilst behavioural biometrics are concerned with our dynamic actions and usage as a unique characteristic [8].

Keystroke dynamics are one of the most widely accepted forms of behavioural biometric currently in use. These describe the unique typing pattern of a user and are collected from keystroke data which results in two measurements; dwell and flight time.

Many different values are calculated from raw keystroke data for analysis purposes. Two of the most common are dwell time (the time between key press and key release) and flight time (the time between the initial key press and pressing the subsequent key) [12]. The interaction of these two values form the user's typing 'rhythm'.

A key consideration when using keystroke dynamics is the split of the letters or text. These are classified as n-grams which can be defined as a contiguous sequence of n items. The most common n-grams are those of uni-grams, bi-grams and tri-grams with these being a length of 1, 2, or 3 letters respectively. Bi-grams are often seen as more of a robust method of identification as they are less effected by environmental factors [1]. The most commonly used n-gram is the bi-gram, and can be found in various experiments.

Mouse dynamics is described as the way in which a user interacts with their system through the mouse. As with keystroke dynamics, there are common measurements which are extracted, which are based around the xy co-ordinates and time [11].

"Soft" biometrics are features which are not uniquely identifying on their own, but can aid in identification [7].

Research into using keystroke dynamics to determine soft biometric features was first conducted by Giot and Rosenberger [4], who were able to predict gender with 91% accuracy. Research conducted by Idrus et al [5] then found accuracy scores of between 80-90% for recognising number of hands used, handedness, age and gender and identification. Soft biometrics are therefore a viable inclusion, and one which can be predicted by utilising keystroke data. This was improved upon by the same authors in the following year [6].

With regards to discovering soft biometrics through mouse dynamics, research has found the gender of a user of a webpage to a success rate of 70%, with the age of a user (either older or younger than 33) with a 90% success rate [9]. Yamauchi and Bowman [18] investigated the identification of gender and emotions of a user while using mouse dynamics. This shows as a 72-74% accuracy for emotion, depending on which emotion was being predicted, and a 61-65% accuracy on gender.

The combination of keystroke dynamics and mouse dynamics is a novel technique which has been researched only partially. Pentel's [15] findings show that mouse was the most accurate compared to keyboard, with scores on identifying gender at 73% for keyboard and 94% for mouse. A combination of keystroke and mouse has also been found to provide the best results when attempting continuous authentication, as it proved difficult to spoof both measurements simultaneously [11].

The studies above highlight the promising potential accuracy of keystroke and mouse dynamics when combined; and also their individual ability to predict some soft biometric features. We present in this paper our preliminary investigation into this concept.

3 Methodology

For this study, we devised a platform which would allow us to gather keystroke and mouse dynamic data through a series of tasks. For the mouse data, participants were asked to click the centre of a single 'crosshairs'. When they clicked within 100px of the centre of the target, the next was shown. This task was similar to that devised by Van et al [17]. Additionally, mouse data was gathered at the point that we collected demographic information. For the keystrokes tasks, participants were first asked to copy a passage from Bram Stoker's Dracula, and then asked to describe the plot of their favourite film. These provided some data with the same expected results for all participants (the fixed text and the mouse tasks), and some which simulated more realistic use (free prompt and mouse data gathered elsewhere). The free collection is crucial should keystroke or mouse dynamics be integrated into a continuous authentication system [16].

3.1 Mouse Dynamics

We captured the mouse data using the p5.js library [13] which allowed us to capture more real-time information in a more intuitive format than basic JavaScript mouse event listeners allowed. The xy co-ordinates and time were logged whenever the mouse moved, the left click button was pressed, or when the right click button was released. Additionally, it was logged whether the clicks were on or off target.

From this data, we extracted a number of features which were mostly selected due to them being common in existing literature. When we combined these features within all of the relevant tasks this created a total of 55 mouse features, as for the features for all mouse data had 4 instances per participant, and the features for mouse tasks only had 3 instances per participant.

3.2 Keystroke Dynamics

Keystroke data was captured using JavaScript keyUp and keyDown event listeners. On any event, the timecode, keycode, and type of event (up or down). From this data, we processed it into keystroke events (Key, Time of Press, Time of Release).

From this data, we extracted a number of features, again chosen from the literature. When we combined these features within all of the relevant tasks this created a total of 186 keystroke features.

3.3 Machine Learning

For classification of the models, we used machine learning as it has generally been found to be more accurate than statistical methods [3]. For this we used the open-source Python library scikit-learn [14]. In order to examine whether combining the two biometrics increased accuracy, we used 5 classifiers: Decision Trees, Random Forest, Gaussian Naive-Bayes, SVM, and K Nearest Neighbours. We have included the results for 2 different sets of parameters for Decision Trees within this paper.

We randomly split our data with 95% in the training set and the remaining 5% in the test set. The data was then resampled 100 times. We additionally performed random undersampling, using the imbalanced-learn library [10].

In addition to undersampling, we also attempted feature selection to improve the accuracy. To do this we used the scikit-learn 'SelectKBest' function, to select the 'k' best features within the dataset, with k set to all, 100, and 150, to examine how reducing the number of features might improve the accuracy. These values for K were selected in order to present a broad range of potential feature numbers, without removing so many features that the models would become imprecise. Due to the smaller number of features, we only used k=all for the purely mouse dynamics experiments.

4 Results

We recruited 240 participants who completed our study. Of these 225 were right handed, 15 left. 120 identified as female, 119 as male, and 1 as 'other', and there was an even distribution of ages, which we split into 6 bins. We cut the participant who identified as 'other' from the sample when examining gender, due to the severe class imbalance. We additionally collected data about the number of hours participants spent on electronic devices in a day, which was again split into 6 bins.

As can be seen in Table 1, the results we achieved varied massively in accuracy between both the different demographics and different classifiers. However; consistently for all the demographics we studied, a combination of keystroke and mouse dynamics provided the most accurate result across all classifiers.

For all demographics, the least consistently accurate classifier was SVM, whilst the most consistently accurate classifier was Random Forest.

		Keystrokes				Mouse				Combo			
		Gender	Hand	Age	EH	Gender	Hand	Age	EH	Gender	Hand	Age	EH
Random Forest (100)	K=all	0.57417	0.61	0.20091	0.06	0.596	0.55000	0.22714	0.11	0.64333	0.58	0.25364	0.085
	K=100	0.60333	0.56	0.2	0.165	N/A	N/A	N/A	N/A	0.68333	0.64	0.26909	0.165
	K=150	0.5775	0.53	0.20909	0.12	N/A	N/A	N/A	N/A	0.6625	0.575	0.26333	0.13
Decision Trees (3, 5)	K=all	0.50417	0.58	0.15545	0.15	0.57667	0.57333	0.23071	0.1	0.52	0.625	0.19727	0.21
	K=100	0.52083	0.575	0.17545	0.19	N/A	N/A	N/A	N/A	0.57917	0.675	0.23273	0.2
	K=150	0.50583	0.58	0.16727	0.175	N/A	N/A	N/A	N/A	0.54167	0.67	0.21	0.16
Decision Trees (10, 3)	K=all	0.5725	0.69	0.14636	0.095	0.57267	0.52333	0.2214	0.16	0.54333	0.805	0.22	0.05
	K=100	0.504167	0.66	0.17909	0.085	N/A	N/A	N/A	N/A	0.54083	0.805	0.21909	0.05
	K=150	0.51167	0.735	0.16	0.1	N/A	N/A	N/A	N/A	0.55917	0.74	0.21273	0.07
SVM	K=all	0.43583	0.315	0.14	0.02	0.50533	0.38	0.13714	0	0.44	0.3	0.14091	0.01
	K=100	0.54583	0.29	0.15364	0.025	N/A	N/A	N/A	N/A	0.6	0.305	0.17	0.025
	K=150	0.515	0.29	0.14727	0.02	N/A	N/A	N/A	N/A	0.50667	0.31	0.14545	0.015
Gaussian Naive Bayes	K=all	0.495	0.26	0.18273	0.1	0.542	0.63667	0.25643	0.12667	0.46167	0.315	0.19818	0.1
	K=100	0.53667	0.425	0.23909	0.11	N/A	N/A	N/A	N/A	0.53083	0.535	0.27636	0.22
	K=150	0.502	0.325	0.17818	0.08	N/A	N/A	N/A	N/A	0.525	0.42	0.24455	0.145
KNN	K=all	0.49833	0.6	0.11455	0.11	0.58267	0.41	0.17857	0.13998	0.51917	0.58	0.138181	0.16
	K=100	0.52417	0.55	0.2	0.13	N/A	N/A	N/A	N/A	0.5925	0.08	0.19818	0.115
	K=150	0.54083	0.535	0.11818	0.08	N/A	N/A	N/A	N/A	0.58	0.555	0.18273	0.175

Table 1. All Accuracy Scores, with the highest for each demographic highlighted.

4.1 Gender

For gender, we achieved a maximum accuracy of 0.68333, using the random forest classifier. Gender was the most consistently accurately predicted demographic, with accuracy only falling below 0.5 on 4 occasions.

These results were worse than previous studies involving both keystroke dynamics and mouse dynamics (such as those in [5] and [18]), however, this is not surprising as studies due to the difference in data collection methodology, as our data collection was created to mimic an authentication system with a single data capture, whilst many previous studies concerned more continuous data collection in a more natural setting, with more data collection.

4.2 Handedness

With handedness we achieved our highest accuracy of 0.805, using decision trees as the classifier. In stark contrast to gender, handedness had the largest range of accuracy scores, with the lowest at 0.26.

With the results for handedness, it is worth noting that the sample size (after undersampling) was significantly smaller than the other demographics tested (30 total) owing to a small sample of left-handed people. This is likely to be a large factor in the range of results, as one failed prediction massively alters the overall accuracy score.

4.3 Age

We achieved very low accuracy with both age and electronic hours. For age, the most accurate classifier (Gaussian Naive-Bayes with 100 features) achieved an accuracy score of 0.27636. This is slightly better than random chance.

Other studies frequently consider age in a binary form (often under or over 30). We chose to additionally complete classification in this form. This massively increased our accuracy, with the highest accuracy being 0.69222 using a combination of keystroke and mouse dynamics with the Random Forest classifier.

4.4 Hours Spent on Electronic Devices

As with age, our accuracy for hours spent on electronic devices was also exceptionally low, with a maximum accuracy of 0.22. The majority of classifiers produced exceptionally low accuracy scores. This suggests that prediction of time spent on such devices is difficult from keystroke and mouse dynamics, potentially because there is not a link between the two.

4.5 Feature Selection

In addition to the above analysis, we also considered which features had been removed during our feature selection stage, to determine if keystroke or mouse dynamics had any meaningful weight on the accuracy. As can be seen in Table 2,

the mouse dynamic features always made up a smaller percentage of the selected features, as it so be expected with them making up just over 22% of the total features possible. This data suggest that mouse features are more discriminatory, as they always made up a larger percentage of features when 100 are selected, and reducing in percentage when the number of features increases to 150. This suggests strongly that the most prejudicial features in the dataset are those arising from mouse dynamics.

It is also important to note that we collected mouse dynamic data in 4 separate parts of the study, compared to the 2 where we collected keystroke data, and this may affect how influential the features are.

	K=100		K=150	
	Mouse Features	Keystrokes Features	Mouse Features	Keystrokes Features
Gender	29%	71%	26%	74%
Handedness	39%	61%	30.67%	69.33%
Age	34%	66%	26.67%	73.33%
Electronic Hours	29%	71%	25.33%	74.67%

Table 2. Percentage split of features

5 Conclusion

The preliminary results of our study show that using a combination of mouse and keystroke dynamic features is more effective at predicting soft biometric features than using either in isolation (see Table 1). As can be seen from the table, the combined accuracy scores were better than the individual (mouse or keystroke in isolation) accuracy scores, for each machine learning method.

In addition to the results in Table 1, we can see from Table 2 that both mouse dynamic and keystroke features were selected as the most influential, showing the value of using them in combination.

Future work will need to focus on a few areas. Firstly, some further tuning of the hyperparamaters of the machine learning methods is necessary to improve accuracy. Additionally, some work with classifiers designed for imbalanced classes might be useful to avoid the need for undersampling, and make the system more reliable in real-world scenarios. Finally, some further investigation into which features were the most prejudicial would be interesting.

References

1. Bergadano, F., Gunetti, D., Picardi, C.: User authentication through keystroke dynamics. ACM Transactions on Information and System Security (TISSEC) **5**(4), 367–397 (2002)

2. Bhatnagar, M., Jain, R.K., Khairnar, N.S.: A survey on behavioral biometric techniques: Mouse vs keyboard dynamics. Int. J. Comput. Appl **975**, 8887 (2013)
3. Giot, R., El-Abed, M., Rosenberger, C.: Keystroke dynamics with low constraints svm based passphrase enrollment. In: Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on. pp. 1–6. IEEE (2009)
4. Giot, R., Rosenberger, C.: A new soft biometric approach for keystroke dynamics based on gender recognition. International Journal of Information Technology and Management **11**(1-2), 35–49 (2012)
5. Idrus, S.Z.S., Cherrier, E., Rosenberger, C., Bours, P.: Soft biometrics for keystroke dynamics. In: International Conference Image Analysis and Recognition. pp. 11–18. Springer (2013)
6. Idrus, S.Z.S., Cherrier, E., Rosenberger, C., Bours, P.: Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords. Computers & Security **45**, 147–155 (2014)
7. Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: International conference on biometric authentication. pp. 731–738. Springer (2004)
8. Jain, A.K., Flynn, P., Ross, A.A.: Handbook of biometrics. Springer Science & Business Media (2007)
9. Kratky, P., Chuda, D.: Estimating gender and age of web page visitors from the way they use their mouse. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 61–62 (2016)
10. Lemaître, G., Nogueira, F., Aridas, C.K.: Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. Journal of Machine Learning Research **18**(17), 1–5 (2017), http://jmlr.org/papers/v18/16-365.html
11. Mondal, S., Bours, P.: A study on continuous authentication using a combination of keystroke and mouse biometrics. Neurocomputing **230**, 1–22 (2017)
12. Moskovitch, R., Feher, C., Messerman, A., Kirschnick, N., Mustafic, T., Camtepe, A., Lohlein, B., Heister, U., Moller, S., Rokach, L., et al.: Identity theft, computers and behavioral biometrics. In: 2009 IEEE International Conference on Intelligence and Security Informatics. pp. 155–160. IEEE (2009)
13. p5.js: p5js.org (2014)
14. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
15. Pentel, A.: Predicting age and gender by keystroke dynamics and mouse patterns. In: Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization. pp. 381–385 (2017)
16. Sim, T., Janakiraman, R.: Are digraphs good for free-text keystroke dynamics? In: 2007 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–6. IEEE (2007)
17. Van Balen, N., Ball, C., Wang, H.: Analysis of targeted mouse movements for gender classification. EAI Endorsed Transactions on Security and Safety **4**(11) (2017)
18. Yamauchi, T., Bowman, C.: Mining cursor motions to find the gender, experience, and feelings of computer users. In: 2014 IEEE International Conference on Data Mining Workshop. pp. 221–230. IEEE (2014)