Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State

Avar Pentel

Tallinn University, Institute of Informatics, Tallinn, Estonia

pentel@tlu.ee

Abstract. This paper describes an unobtrusive method for user confusion detection by monitoring mouse movements. A special computer game was designed to collect mouse logs. Users' self-reports and statistical measures were used in order to identify the states of confusion. Mouse movement's rate, full path length to shortest path length ratio, changes in directions and speed were used as features in the training dataset. Support Vector Machines, Logistic Regression, C4.5 and Random Forest were used to build classification models. Models generated by Support Vector Machine yield to best classification results with f-score 0.946.

Keywords: confusion detection, behavioral biometrics, mouse dynamics.

1   Introduction

The ability to recognize, interpret and express emotions plays a key role in human communication and increasingly in HCI. In the context of learning systems, the ability to detect user emotional states, gives promising applications to adaptive recommendations, adaptive interfaces, etc. Usually special equipment is used for emotion detection: electroencephalogram, skin conductance, blood volume pressure [1,2] or gaze and facial data [3,4]. But when it goes to real life application, we can relay no more, than unobtrusive standard computer inputs like mouse or keyboard.

The theory of "embodied cognition" [5] gives a theoretical framework studying mouse movements in order to predict mental states. Barsalou suggests that this bi-directional relationship between mental states and bodily states emerges because the core of social and cognitive information processing lies in the simulation of original information [6]. There are some studies [7,8,9,10] about mouse movement and emotions, which all suggest a link between mouse movement and emotions. Yet, most of these studies are conducted with relatively small samples. Secondly, all these studies are dependent on the specific context of an experiment, and general link between emotions and mouse movements is not investigated.

In the current study, we aim to find a link between confusion and mouse movements and try to avoid both of previously mentioned shortcomings by using larger sample, and avoiding specific context in our experiment.

2 Methodology

2.1 Data Collection Procedure and Sample

A simple computer game was built to collect user mouse data. The idea of the game come from Christmas Calendar chocolate boxes, where the chocolates are hidden behind numbered doors. There are usually numbers from 1 to 24, and in order to make the right door harder to find, numbers are randomly arranged and they look differently. Similarly, we designed a game, which fills screen with randomly arranged buttons labeled with numbers 1 to 24. All buttons are different size and color (Fig.1). User task is to click on all buttons in the right order as fast as possible. To keep up the motivation, the game was installed in school computer class as a part of login system, i.e. in order to log in users were forced to play this game. There was also an option to play it many times. It was also publicly announced, that best performers would be awarded. For every game session mouse activity (movements and clicks) was logged.
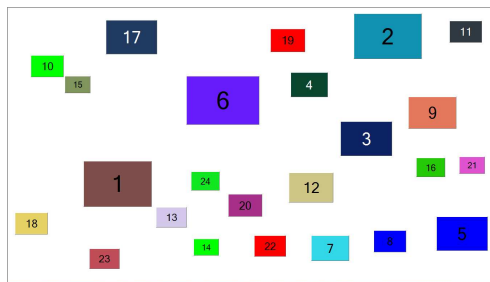

Fig. 1. A Christmas Calendar game built for data collection. The user has to click as fast as possible on all buttons in the right order

Our logging procedure was an event based, which means that mouse position was not recorded in fixed interval, but only if difference in position of mouse occurred. In our case this difference of position was set to 30 pixels. Our mouse logs consisted triples of x and y coordinates and timestamp. We recorded data from 516 game sessions played by 262 individual users. As each game session consisted of 24 searching tasks (to find next number), we had all together 12384 comparable records, each of them presenting mouse movement logs between two button clicks.

2.2 Labeling Data with Emotional state

We also interviewed selected participants (N = 44) right after the game. Reviewing together the whole game session again, we asked to describe the emotions during the game. Initially we asked users to position his/her emotions on Russell's circular model [11], but pre testing revealed, that in the current set of the experiment, users were only able to describe two categories of emotions - the state of confusion and the state of content. Therefore we continued to collect self-report data on a 7 point Likert scale where 1 = content, and 7 = confused. While users were not able to specify the

exact time when the state of confusion began or end, we divided the game session to 24 separate searching tasks, and linked those emotion feedback data to a whole task. All together we got 44 x 24 = 1056 tasks labeled with emotion data.

It is intuitively clear, that in such circumstances, confusion and target finding speed are related. While target finding speed differs individually, all these finding scores were standardized session-wise, and then Pearson correlation with confusion self-report data were found. As expected, there was significant correlation between confusion and standardized finding time (r = 0.86). Also, all tasks associated with confusion, had standardized finding speed half standard deviation below the mean, and those associated with a feeling of content, half standard deviations above the mean. While our interviews covered only less than 10% of all game sessions, we extended this relation to all other game sessions too.

We suppose, that very quick results may not include confusion at all, i.e. user is aware from the beginning about the location of the target. But in order to minimize possible confusion, which may be present in the beginning of each task, we divided finding time to the half and used only the last half of log data as characterizing non-confusion. Similarly, it is obvious, that in tasks that were characterized as confusing, the state of confusion does not cover the whole time between two button clicks. Obviously, confusion must end in some moment, when user notices the next button. It is reasonable to suppose that confusion ends somewhere in the second half of the searching process. Therefore, we split each of these slower result logs to the half and used only the first half of searching task as characterization of confusion (Fig. 2.).
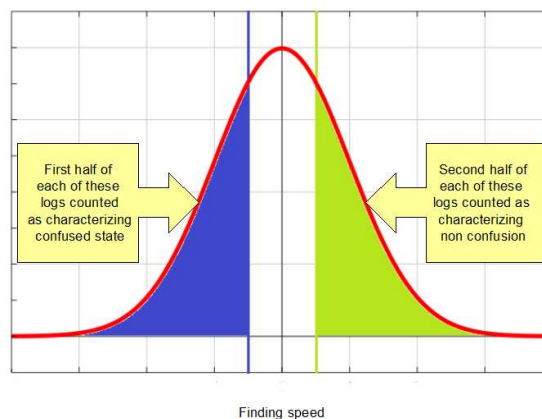

Fig. 2. Separation of mouse logs representing state of confusion and non-confusion.

Out of these two subsets we excluded repeated sessions by the same users, and extreme results. From the remaining data we created balanced training dataset of 2282 records.

2.3 Features

In the current study, we extracted 33 features based on distance, speed, direction and direction change angles (Table 1.). Feature selection procedure with Chi Squared attribute evaluation and SVM attribute evaluation revealed, that strongest features were those of speed based and those of based on relations of shortest distance and actual distance. Best models with those attributes yield to F-score 0.96 width SVM and Logistic regression.

Table 1. Features.

Type	Feature	Explanation
Distance*	Precision	Shortest distance between two button clicks and actual mouse path length ratio.
Speed**	Speed	Actual mouse path length between two button clicks divided by task completion time.
	AdjSpeed	Actual mouse path length between two button clicks divided by shortest path, and then divided by task completion time.
Direction	DirectionX	Number of mouse movements in particular direction. We divided movement directions to 8 distinctive segments as north, northeast, east, etc. We counted all movements in particular direction segment, and divided to all movements.
Direction changes	TurnA Turn10, Turn20,... Turn 180 TurnA+	Mouse movements' path was recorded as consecutive straight lines of 30px length. We measured each angle between two consecutive movements and extracted 18 features representing turns from 0 to 180 degrees by 10-degree step. Counted results were normalized by whole number of movements. All turns greater than angle A (A counted by 45-degree step).

* Excluded in training feature set of the models titled as "target unknown" in Table 2.
** Excluded in all training feature sets.

For our final model we had to exclude features that were related to speed, because speed was previously used by us for associating tasks with emotional states. Without speed related features, models F-score dropped from 0.96 to 0.946.

As our goal was to identify confusion patterns without knowing the real target, we also excluded the feature that was calculated by using information about shortest distance. All reminded features were based on movement direction and direction changes. Direction based features were number of movements on specific direction divided by mouse path length. Direction changes were measured as the angle between previous and next movement. Within these features strongest features were direction changes closer to 180 degrees, more than 135 degrees and between 160 and 170 degrees.

2.4 Machine Learning Algorithms and Technology

For classification we tested four popular machine-learning algorithms: Logistic regression, Support Vector Machine, Random Forest, and C4.5. Motivation of choosing those algorithms is based on literature [12,13]. The suitability of listed algorithms for

given data types and for given binary classification task was also taken in to account. In our task we used Java implementations of listed algorithms that are available in freeware data analysis package Weka [14].

For evaluation, we used 10 fold cross validation. We partitioned our data into 10 even sized and random parts, and then using one part for validation and another 9 as training dataset. We did so 10 times and then averaged validation results.

3 Results

As mentioned before, when excluding all speed-based features, our SVM model with standardized data yield to F-score 0.946. When excluding all distance-based features, results dropped considerably, but all our classifiers still yield to F-scores over 0.8. In following table (Table 2.) are presented results of different classifiers generated with features that are calculated using data about known target (i.e. the shortest path) and without these features.

Table 2. The results of the models trained with different feature sets

Model	Target known			Target unknown		
	Accuracy	*F-score*	*ROC*	*Accuracy*	*F-score*	*ROC*
SVM (standardized)	94.61%	0.946	0.946	82.38%	0.824	0.825
Logistic Regression	93.49%	0.935	0.978	82.72%	0.827	0.889
Random Forest	92.07%	0.921	0.971	84.47%	0.845	0.825
C4.5	91.96%	0.919	0.937	83.59%	0.835	0.836

4 Discussion and Conclusion

Simple feature set of directions, direction changes and relations between actual and shortest distance proved to be useful in classification confused and non-confused user. As we can see from Table 1, knowing the target makes predictions better, but even without knowing the target, frequent direction changes in mouse movement, are still good predictors of confusion. This might be an indirect confirmation to studies about the correlation between gaze and mouse movements.

However, we have to address the limitations of such set of experiment. Depending on the tasks and page layout, user mouse movements might differ considerably. Our results are applicable in situations, where users have to find something particular on unfamiliar (web) environment, in set of menus, links or graphical elements. But our approach might not work in web page considered for reading. For example, while somebody is used to fallow line with mouse cursor while reading the text, the mouse logs will show frequent changes in directions, which in by our model will predict confusion. Therefore more study is needed in different types of environments.

References

1. G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion Assessment : Arousal Evaluation Using EEG's and Peripheral Physiological Signals," in Proc. Int. Workshop on Multimedia Content Representation, Classification and Security, pp. 530-537 (2006)
2. E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, "A user-independent real-time emotion recognition system for software agents in domestic environments," Engineering Applications of Artificial Intelligence, vol. 20, no. 3, pp. 337–345 (2007)
3. Happy, S.L. et al. Automated Alertness and Emotion Detection for Empathic Feedback During E-Learning, IEEE Digital Library (2013)
4. Jaques, N. et al. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System, Lecture Notes in Computer Science Volume 8474, pp 29-38 (2014)
5. P. M. Niedenthal, "Embodying emotion," Science, vol. 316, pp. 1002-1005, (2007)
6. L. W. Barsalou, "Grounded cognition," Annual Review of Psychology, vol. 59, pp. 617-645, (2008)
7. J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard, "Frustrating the user on purpose: a step toward building an affective computer," Interacting with computers, vol. 14, pp. 93-118, (2002)
8. Zimmermann, P. S. Guttormsen, B. Danuser, and P. Gomez, "Affective computing - a rationale for measuring mood with mouse and keyboard," International journal of occupational safety and ergonomics, vol. 9, pp. 539-551, (2003)
9. Zimmermann, P. "Beyond Usability–Measuring Aspects of User Experience," Doctoral Thesis (2008)
10. Maehr, W. eMotion: Estimation of User's Emotional State by Mouse Motions: VDM Verlag, (2008)
11. Russell, J. A. 1980. A Circumplex Model of Affect. Journal of Personality and Social Psychology. Vol. 39, No. 6. 1161-1178 (1980)
12. Wu, X. et al. Top 10 algorithms in data mining. Knowledge and Information Systems. vol 14, 1–37. Springer (2008)
13. Mihaescu, M. C. Applied Intelligent Data Analysis: Algorithms for Information Retrieval and Educational Data Mining, pp. 64-111. Zip publishing, Columbus, Ohio (2013)
14. Weka. Weka 3: Data Mining Software in Java. Machine Learning Group at the University of Waikato. http://www.cs.waikato.ac.nz/ml/weka/