A review of emotion recognition methods based on keystroke dynamics and mouse movements

A. Kołakowska
Gdansk University of Technology, Gdańsk, Poland,
agatakol@eti.pg.gda.pl

Abstract. **The paper describes the approach based on using standard input devices, such as keyboard and mouse, as sources of data for the recognition of users' emotional states. A number of systems applying this idea have been presented focusing on three categories of research problems, i.e. collecting and labeling training data, extracting features and training classifiers of emotions. Moreover the advantages and examples of combining standard input devices with other sources of information on human emotions have been also described. Eventually some conclusions from the review have been drawn.**

Keywords: **emotion recognition, keystroke dynamics, mouse movements**

I. Introduction

Recognizing emotions of computer users became an intensively investigated research area in recent years. A great many methods have been developed and explored in order to identify how a user feels. These methods are based on different data sources, i.e. visual (facial expression, gestures) [1, 2], audio (speech, voice) [3], textual (semantics), physiological (heart rate, temperature, skin conductance etc) [4], input devices (keyboard, mouse, touch-screen) [5-7] and even on a combination of them [3, 8, 9]. Affect-sensitive systems may not only recognize users' emotions, but also react accordingly to the detected states. The popularity of studies on affective computing arises from the possibilities of applying these methods in real-life domains, such as for example intelligent tutoring systems (ITS) [10], adaptive interfaces [1], gaming, software engineering [11], designing web-pages and others.

This paper focuses on the affective computing methods based on data coming from standard input devices, which enable systems to recognize users' emotional states. Section II presents keyboard and mouse as the sources of biometric characteristics. Section III describes methods using this type of data, including data gathering, feature extraction, training classifiers and recognizing emotions. Section IV presents a few multimodal approaches and section V gives some conclusions.

II. Keystroke dynamics and mouse movements as biometric features

Analysis of keystroke dynamics and mouse movements is not a novel idea. These are common biometric techniques used for ensuring the security of computer systems. There is a large group of methods, designed to verify one's identity, based on human characteristics. Biometric features may be either physiological (face, palm, fingerprint, iris) or behavioral (voice, handwritten signatures, keystroke dynamics, mouse movements). Most security systems are based on physiological features, because of their stability along time. The difficulty with behavioral features is that the samples of one user may vary strongly every time they are recorded. One's typing rhythm or pointer movements strongly depend on the type of hardware and software used, and even on the time when they are recorded. The advantage of the keystroke dynamics or mouse movements is that they are natural type of biometrics which do not require any special hardware. Moreover, they are not as intrusive as some other methods. It is possible to record the keyboard and mouse parameters during the usual computer usage.

A number of research on keystroke dynamics as biometrics have been described. Some of them are attempts to authenticate users on the base of a fixed text [12, 13]; whereas others analyze free texts [14, 15]. All of them incorporate features, which are based on time durations between the keystrokes, inter-key strokes and dwell times, which is the time a key is pressed down, typing speed, frequency of using backspace key etc. [16].

Using mouse movements as biometrics is less common than the keystrokes, but it is also investigated [17]. The methods of this group analyze mouse characteristics such as mouse speed, acceleration, direction, number of clicks etc. A combination of both keystroke and mouse movements has been also explored and applied to detect intruders [18].

III. Emotional states vs. keyboard and mouse

Creating an emotion recognition system is a challenging task. It requires solving dozens of problems, which may be grouped into three main categories: data gathering, data representation and training the system. The following subsections describe different solutions proposed for these three stages.

A. Collecting and labeling data

As it has been already mentioned, collecting data from keyboard and mouse is not intrusive in contrast to other devices, such as for example cameras or sensors monitoring physiological parameters. In spite of this it is still quite difficult, because of the amounts of data which should be gathered. Moreover it is also necessary to label the recorded sequences of measurements, i.e. assign them labels of

978-1-4673-5637-4/13/$31.00 ©2013 IEEE

proper emotional states. Thus it is very important to precisely think over the scenario of an experiment.

The authors of [5] described an experiment in which participants, after being affected by video clips, were supposed to perform an online-shopping task. They labeled their affective state via a Self-Assessment-Manikin questionnaire, which presents emotions in a graphical way [19]. Five emotional states were identified: PVHA, PVLA, NVHA, NVLA and nVnA (P-positive, N-negative, H-high, L-low, n-neutral, V-valence, A-arousal). Moreover a few physiological parameters were measured: hart rate, skin conductance, respiration and corrugators activity, but they were not analyzed in the experiment.

A research on identifying emotional states using keystroke dynamics in a real-life application has been presented in [6]. The authors gathered keystroke data from 12 users during their usual daily computer activities, such as using word processor, mail or messaging applications. The users were interrupted from time to time, depending on their activity, by presenting their recently typed text and asking to fill in an emotional state questionnaire. Then they were also asked to retype a fixed text. The questionnaire contained 15 questions, each for one of the following emotional states: anger, boredom, confidence, distraction, excitement, focused, frustration, happiness, hesitance, nervousness, overwhelmed, relaxation, sadness, stress, tired. The answers were given in a 5-point Likert scale (strongly disagree, disagree, neither disagree nor agree, agree, strongly agree) [20]. The advantage of this approach was no need to induce any emotional states and no limits on the types of emotions, which therefore did not have to be predefined. However the main drawback was no control on the states tested and the number of samples collected per state.

In [21] an application of classifying emotions in an intelligent tutoring system has been presented. Twenty participants of the described experiment were given questions on computer science. Their answers were evaluated depending on the presence of keywords from possible correct answers. Depending on the result and the number of already performed trials they were either praised, asked to try again, elaborate the answer or simply informed about the mistake. Moreover after each answer the participants were asked to describe their emotions by filling a short questionnaire which asked them to agree or disagree with five statements (confusion, boredom, frustration, delight, neutral) using a 5-point Likert scale. The evaluation was made not only by the participants, but also by two judges trained in emotional intelligence, who watched the videos recorded during experiments. It resulted in three datasets with different labeling. Further experiments were performed and compared for these three sets. The response quality, correctness and spelling mistakes were evaluated only by human judges.

Keystroke rhythm may be analyzed not only on a PC or laptop, but also other devices, such as mobile phones. An example application of this type has been described in [7]. The authors created an application, which after installing on a smartphone, gathered various sensor data from it. The phone user was asked to write short messages reporting his emotional state, whenever he felt a specific emotion. Seven emotional states were predefined: happiness, surprise, anger, disgust, sadness, fear and neutral.

The aim of the experiment presented in [22] was to answer the question whether it is possible to detect stress on the base of keystroke and linguistic features. The data was collected from 24 participants during a few sessions. The first four sessions were treated as a baseline and control conditions. During each of them a participant entered three free texts on any topic and of any length. The next typing sessions were completed after performing some stress tasks. Two types of stress were taken into account. The first two tasks (mental multiplication and three-back number recall) were supposed to induce cognitive stress. Two other tasks were bringing about physical stress by making users walk on a treadmill or use a resistance band. After each task the participants had to enter a free text in the same way as during the control sessions. Moreover they were asked to describe the level of stress using an 11-point Likert scale.

Another example of research on predicting mood of an ITS student has been presented in [23]. The authors gathered samples from 124 participants who were trained in first aid using a multimedia presentation. Before the training the participants gave some biographical information, such as age, gender, first aid experience etc. Their knowledge and mood were assessed before and after the training. To describe mood the Self Assessment Manikin questionnaire was used. Apart from knowledge the participants' performance was also assessed after the training, i.e. their ability to apply knowledge making proper decisions and actions.

In the experimental study described in [24] 41 participants were supposed to retype a few fixed texts inducing different emotions. The texts were recorded at different times to enable capturing keystroke information under different emotional states. After typing the participants described their emotional state as belonging to one of three categories: positive, negative, neutral.

Another approach has been presented in [25], where pressure sensor keyboards were used, so not only traditional keystroke parameters were measured, but also pressure sequences. 50 participants listened to a story for each of the following six emotions: anger, fear, happiness, sadness, surprise and neutral. The reactions to the story was reflected by typing a text 10 times for each story. In this way 3000 labeled samples, all of the same length, were gathered.

The author of a research on the influence of emotions on mouse movements presented in [26] chose to induce emotional states by two videos – a happy and a sad one. The information about emotional states was gathered via an online questionnaire. Filling the questionnaire required using a mouse, so it was possible to record all data from that input device, but during the experiment the participants were not aware of this fact. Three videos were shown to each of 40 participants. The first one was neutral and it was followed by general questions on age, gender, experience etc. and then by an emotion questionnaire. The other two videos were to elicit sadness and happiness respectively. Each video was followed by questions on user's emotional state.

The aim of another mouse movement study described in [10] was to detect boredom while using an online course on computer programming techniques. During the experiment all mouse data was recorded. If the mouse stopped for a time longer than a predefined threshold the student was asked whether or not he was bored. The data was collected from 136 participants.

As it can be seen from the presented experiment scenarios there are a few questions to be answered before starting the study. What kind of emotions will be taken into account? How to induce the emotional states? How to label the samples? How much data should be gathered?

B. Extracting features

After collecting the data, it is necessary to represent the samples in a way recognizable by the algorithms which are going to be used. The samples are presented as feature vectors. The dimension of the vectors, i.e. the number of features, and the type of the features is determined by the type of data recorded.

Keystroke features may be divided into the timing and frequency characteristics. Timing features may measure the duration of single keystrokes or sequences of keystrokes (digraphs, trigraphs etc.), the times between subsequent keystrokes, the times between depressing a key and pressing the next one and the typing speed. Frequency features measure how often keys are pressed. Usually it applies to special keys, such as backspace, delete, caps lock, numpad keys, punctuation keys. These statistics may indicate for example the way a user corrects mistakes or user's care of how the text is written. As these behaviors may be affected by emotions, keystroke parameters might be indicators of some emotional states.

Mouse features may be divided into those which describe the way of moving a pointer and those connected with clicking or scrolling. Mouse movement parameters describe the lengths of paths, their direction, path length to direct distance ratio, speed, acceleration. The second group contains the frequency of different types of clicks and different type of actions, i.e. drag and drop, point and click etc. Various combinations of these parameters may measure for example the smoothness of the pointer paths, which in turn may show user's confidence.

It is obvious that the values of keystrokes and mouse parameters taken from one user strongly depend on the type of software used. Keystroke frequencies while using an Internet browser are incomparable to those when a word processor is used. Mouse measurements taken via an interface of a graphical program are completely different than those while using a word processor. That is why it is extremely difficult to build an inference system on the base of this type of data without accounting the application used.

The typical keyboard and mouse features are often accompanied by some additional characteristics as it can be seen from the following research examples.

During the experiment presented in [5] all mouse and keyboard data were recorded giving 64 features. Some of them were: the total number of mouse clicks, single mouse clicks (multiple clicks were counted as one click), total distance of mouse pointer, mouse speed, time between pressing and releasing a mouse button, number of pauses in mouse movements, median distance of a single mouse movement, mouse acceleration, angle and direction of mouse movements, number of keystrokes, median length of a keystroke etc. The initial set of 64 parameters was then reduced to 16 by eliminating those showing the highest correlation.

More than 30 features were extracted from the gathered raw data during the experiments on usual daily computer activities presented in [6]. These were: duration times for single keys, digraphs and trigraphs; latency times for digraphs and trigraphs; times between pressing two successive keys in graphs; number of events in graphs; number of characters, numbers, punctuation marks, uppercase characters. After removing outliers, performing feature selection and an under-sampling method to avoid strongly unbalanced class distribution, the training set was ready to train the classifier.

During a user's interaction with the intelligent tutoring system described in [21] keystroke data was recorded. The system extracted 18 features divided into the following three groups: timing features (session duration, pause rate), typing features (typing speed rate, key latency, key duration, deletion rate, capitalization rate, spaces per response, punctuation rate, unrelated keys rate), response features (response quality, response correctness, number of words, spelling mistakes, attempts per question). The set of 18 features was then reduced using principal component analysis. Moreover a few additional parameters were taken into account as features: first language, educational level, user level. The interaction features were normalized using the mean value and standard deviation calculated on the base of baseline data taken from the participants before the exact experiment.

In the research on recognizing emotions of mobile phone users [7] 14 features were extracted from the data. Some of them were typical keystroke parameters: typing speed; the frequencies of pressing backspace, enter and special symbols. There were also two features describing the messages sent by the users: maximum text length, erased text length. Moreover the authors defined characteristics typical for the devices used: touch count, long touch count, device shake count, illuminance. Eventually a few additional factors were also considered: discomfort index, location, time and weather. After collecting the data, feature selection based on gain ratio was performed, which lead to choose 10 features showing the strongest correlation with emotions.

Three groups of features were extracted in the study on detecting stress presented in [22]: keystroke timing, keystroke frequencies and word features. The four timing features were: time per keystroke (total input time/ total number of keystrokes), adjusted time per keystroke (time per keystroke excluding pause time), average pause length and pause rate. Frequency features were the total numbers of selected keys per the total number of keystrokes. They were calculated not only for special keys, but also for letters and numbers. Finally there was a group of more than 20 linguistic characteristics, for example the average sentence length, the frequency of using nouns, verbs, modal verbs, cognitive words, words indicating negative affect and many others. The baseline data, gathered before affecting users with stress, were used to estimate the mean values and standard deviations and then the experimental data was normalized using those values.

Eight features were extracted in the study presented in [24]: the mode, standard deviation, variance and range of the typing speed; number of characters typed in 5 second intervals; total time spent on typing; idle time and the number of pressing backspace key.

In the experiments described in [25], where pressure sensor keyboards were used, three groups of features were defined. The first group consisted of five global parameters of a pressure sequence representing each keystroke: mean value, standard deviation, the difference between max and min value, the positive energy center and the negative energy center. The second type of features were the sequences of pressure values of each keystroke. The third group of parameters were typical times between successive keystrokes and the times between depressing a key and pressing the next one in the sequence.

Mouse movement analysis supporting emotion recognition among Internet browser users has been implemented in the system presented in [3]. The features used in that study were the coordinates of cursor and their first and second order derivatives. In the cases of interacting with the application via touch-screen, one more coordinate was added to represent the strength of the touch. In this application pointer movements analysis was accompanied by speech recognition.

A thorough study on the influence of emotions on mouse movements has been described in [26]. The following parameters were analyzed in that research: acceleration, precision, smoothness, uniformity and speed. Acceleration and deceleration were averaged over splitted movements. Precision was measured by the ratio of clicks performed to the clicks needed to hit a target (lower ratio means higher precision), or by the way the targets were hit (hitting targets directly without moving beyond indicates higher precision). Movement smoothness was defined as the number of movement breaks and the uniformity as standard deviation from the average speed. Speed was represented by the average speed of movements and also as the duration of clicks.

An application of mouse movements analysis in an intelligent tutoring system was presented in [23], where only two mouse parameters were analyzed: control selection rate and mouse movement rate.

Much more mouse parameters were used in another ITS described in [10]: average speed, average speed during the last 60 seconds before reporting the emotional state, the average time of mouse inactivity, the number of mouse pauses during the last 60 seconds; horizontal, vertical and diagonal movements to the total number of movements; average speed per direction for eight ranges of direction angles. One more feature, not connected with mouse movements, indicated one of seven types of learning objects (short text, short text with images, medium text with images, long text with images, video, multiple choice, exercise).

Defining a proper set of features is an important stage of designing an affect-aware system, because it influences the performance of recognition algorithms in the sense of accuracy and complexity.

C. Training classifiers and recognizing emotions

Numerous algorithms may be applied to train a classifier of emotional states. The choice depends on the characteristics of training data. First of all the method has to be appropriate for the type of feature values. Most features coming from keyboard and mouse are numerical, but symbolic parameters may also appear. Sometimes the method must be able to cope with both types of features, in such case a decision tree might be a good choice. The second thing is the number of feature vectors. It is not always possible to gather a large enough data set. In such cases some methods should not be taken into account, e.g. methods based on probability estimations. Moreover the number of recognized emotions is important as well. There is a difference between recognizing a few emotions and detecting one state, e.g. stress. In the second case it is possible to implement a two-class approach. In the first case a multiclass classifier may be trained or a number of two-class models, one for each emotion. Such solution is more common as it may be noticed from the examples below.

The authors of [6] chose a decision tree constructed by C4.5 algorithm as a model. For each of 15 emotional states a two-class classifier was trained to recognize whether that state occurred or not. The experiments were performed separately for free and for fixed texts. In the case of fixed texts the best results were achieved for confidence, hesitance, nervousness, relaxation, sadness and tiredness. The recognition accuracy in these cases ranged from 77.4% to 87.8%. Anger and excitement achieved equally high rates, however the data distributions for the two states were highly unbalanced, so these results were not reliable. The classifiers built for free texts did not perform well. The authors suggested that other machine learning algorithms or other features could improve it.

In the research described in [21] the decision process consisted of two stages. The first step included a two-class classifier to discriminate between positive and negative valence. Then other classifiers were trained to recognize the following emotions related to a learning process: confusion, boredom and frustration in the case of negative valence; delight and neutral in the case of positive valence. Different machine learning methods were applied to recognize these emotions: discriminant analysis (linear and quadratic), naive Bayes, k-nearest neighbor, decision trees and artificial neural networks. The average classification accuracy of emotional valence was 69.17%, 63.19%, and 67.45% depending on a dataset (there were three sets labeled by participants and two judges). Classification accuracy for the emotions recognized during the second stage was on average 44.59%, 34.86%, and 43.59% for the three datasets respectively. The best results were obtained for neural networks (82.82%, 72.02% and 77.20% for emotional valence; 53.59%, 45.60% and 53.89% for the five emotional states). The author also analyzed the correlation between typing features and the emotional states. The analysis indicated that being delighted was associated with more careful writing, including the use of punctuation marks, capitalization, deletion, and long correct answers. Being delighted was also associated with short pause rate and few tries. Multiple interactions usually lead to confusion, which was associated with fast and careless writing, long pause rate, long key latency and multiple attempts. Boredom was associated with long session duration, short pause rate and the use of unrelated keys. Frustration was associated with careless writing, which was manifested by lower deletion rate, lower usage of punctuation and capitalization, short responses, incorrect answers and low quality answers. Other consequences of frustration were multiple attempts and a long pause rate.

To classify emotions of phone users [7] Bayesian network was used as a classifier. After building the classifier, the model was continuously updated using new data. Unfortunately the data was gathered from one user only, so general conclusions could not have been made. However the classification accuracy of these preliminary experiments was 67.52% on average and it strongly depended on the type of emotion. The best recognition rate was achieved for happiness, surprise and neutral state.

The authors of [22] applied a number of machine learning methods to detect stress: decision trees, support vector machine, k-nearest neighbors, AdaBoost and neural networks. The experiments were performed on raw and normalized samples. Separate classifiers were built for recognizing either cognitive or physical stress. In the case of raw data the best accuracy was obtained using Ada Boost algorithm (61.5% for cognitive stress and 62.5% for the physical one). In the case of normalized data k-NN performed the best (75%) for the cognitive stress and both SVM and neural networks for the physical (62.5%). The authors managed not only to detect stress but also to find out whether it was cognitive or physical. Moreover the discrimination power of features was evaluated using information gain criterion. In the group of keystroke features a strong relation was shown between the emotional state and the use of backspace, delete, end, arrow keys and also the time per keystroke and pause length. One of the important conclusions was that the number of mistakes made while typing decreased under stress.

In the study described in [23], where only two mouse parameters were analyzed while using an ITS, a statistical analysis was performed. It turned out that there was a correlation between pleasure variable measured after the training task and the control selection rate. Moreover there was a correlation between participants performance in applying knowledge gained during the training and mouse movement rate. The authors indicate that these two behavioral features may be used as predictors of mood and performance respectively.

The authors of [24] applied several classification methods to recognize positive vs. neutral or negative vs. neutral emotional state: linear logistic regression, SVM, neural networks, random tree, C4.5 tree, binary tree. The best performance was obtained using binary tree for negative vs. neutral (89.02%) and C4.5 tree for positive vs. neutral (88.88%). One of the conclusions of the experiment was that the typing speed usually decreases in the case of negative emotional states and vice versa.

The research presented in [25] is an example of implementing separate nearest neighbor classifiers for the users of pressure sensor keyboards. First the emotions were recognized on the base of only one type of features. In the case of global pressure parameters and traditional keystroke features they applied the distance between two feature vectors as the average absolute difference between their coordinates. Whereas in the case of pressure values dynamic time warping was applied to measure the distance between two sequences. Finally the authors proposed a combined measure as a weighed sum of the three distances. The weights were adjusted during experiments. Using the combined distance measure improved the recognition rate, which was 93.4% after averaging over the six emotional states.

The author of [26] did not apply any classification methods, but he performed a thorough statistical analysis, which led to some conclusions. There was a significant correlation found between the arousal and the movement precision and smoothness. Lower arousal was usually connected with more precise and smooth movements. Higher arousal made the users move the mouse faster with higher acceleration values. This lead to less precise movements. In the author’s opinion precision might be a measure of user’s attention, level of focus, tiredness, hastiness etc. Moreover arousal influenced the number of clicks and movement breaks. It was also shown that the subjects moved the mouse more directly to the target and hit it with less clicks after watching the lower arousal film. In contrast, the higher arousal film made the subjects move the mouse with significantly less breaks. Another important observation was that the films did not provoke exactly the same feelings in different subjects. The author claims that the choice of videos is essential when the emotions are to be elicited in this way.

The recognition of boredom investigated in [10] was performed using a decision tree trained with C4.5 algorithm. The experiments were performed for different values of a threshold parameter, which was the mouse idle time, after which a user was asked whether he was bored. In all cases the achieved recognition rate was above 90%.

TABLE 1: EMOTION RECOGNITION BASED ON DATA FROM STANDARD INPUT DEVICES.

Ref.	Emotions analyzed	Eliciting emotions method	Sources of features	Data labeling method	Method	Results	Application
[5]	PVHA, PVLA, NVHA, NVLA, nVnA (P-positive, N-negative, H-high, L-low, n-neutral, V-valence, A-arousal)	video	keystrokes, mouse movements	questionnaire (Self – Assessment – Manikin)	statistical analysis	conclusion: possible discrimination between neutral category and four others	online shopping
[6]	anger, boredom, confidence, distraction, excitement, focused, frustration, happiness, hesitance, nervousness, overwhelmed, relaxation, sadness, stress, tired	none	keystrokes	questionnaire (5-point Likert scale) filled from time to time	C4.5 decision tree (a two-class classifier for each emotion)	classification accuracy: 77.4-87.8% for some emotions (confidence, hesitance, nervousness, relaxation, sadness, tiredness)	usual daily computer activities, e.g. using word processor, mail application
[21]	confusion, boredom and frustration in the case of negative valence; delight and neutral in the case of positive valence	none	keystrokes, quality of responses, biographical data	questionnaire (5-point Likert scale) filled by participants and two judges	two stages: 1 - positive vs. negative valence; 2 - recognizing six emotions; methods tested: discriminant analysis, naive Bayes, k-NN, decision trees, neural networks	best accuracy achieved for neural networks: 72.02-82.82% at the first stage; 45.60-53.89% at the second stage	intelligent tutoring system
[7]	happiness, surprise, anger, disgust, sadness, fear, neutral	none	keystrokes, touch-screen parameters, discomfort index, location, time, weather	messages reporting emotional state	Bayesian network continuously updated with new data	67.52% on average, the best accuracy for happiness, surprise and neutral state	mobile phones
[22]	stress	tasks inducing cognitive and physical stress	keystrokes, language parameters	questionnaire (11-point Likert scale)	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	
[23]	looking for any influence on mood and performance	none	mouse movements	questionnaire (Self – Assessment – Manikin)	statistical analysis	conclusions: there is a correlation between pleasure variable and the control selection rate; correlation between performance in applying gained knowledge and mouse movement rate	intelligent tutoring system
[24]	positive, negative, neutral	fixed texts to be retyped	keystrokes	questionnaire	linear logistic regression, SVM, neural networks, C4.5 decision tree, random tree, binary tree	accuracy 89.02% (negative vs. neutral using binary tree); 88.88% (positive vs. neutral using C4.5 tree) observation: typing speed decreases in the case of negative emotions	
[25]	anger, fear, happiness, sadness, surprise and neutral	stories to be listened	keystrokes, keyboard pressure sequences	labels assigned according to the type of a story listened	k-NN with different distance measures (including DTW) depending on the subsets of features and a combined distance measure	accuracy 93.4% averaged over emotional states using a combined distance measure	
[3]	surprise, joy, anger, fear, disgust, sadness, neutral	simulated Internet browser	mouse movements, touch-screen		combination of distance based classifiers, DTW	recognition accuracy: 76-95% depending on the type of emotion	Internet browser interface

		provoking emotions	parameters, speech		and HMM		
[26]	sadness, happiness, neutral	videos	mouse movements	online questionnaire	statistical analysis	conclusions: lower arousal means more precise and smooth movements and less clicks; higher arousal means higher acceleration, lower precision and less breaks	
[10]	boredom	none	mouse movements, type of learning object	asking students whether they are bored	C4.5 decision tree	more than 90%	intelligent tutoring system

It is no use comparing the recognition rates of the mentioned methods, because they were all tested for completely different sets of data. However some valuable conclusions might be considered while building a new system. Table 1 summarizes the main features of the presented solutions.

I. Multimodal approach

The accuracy of emotion recognition based only on keystrokes and mouse movements is often lower than required in real-life applications. However it may be treated as an additional source of information and together with other data create data sets for reliable emotion recognition systems.

An example of such application has been presented in [3], where the authors analyzed pointer movements performed using either mouse or touch-screen as an additional data source beside speech signal. The final decision on the affective state was made up by averaging the results obtained on the basis of semantic features, signal pitch and energy, and eventually the data coming from mouse. The classification accuracy was increased using this combination and it varied among the emotional states.

In [27] in has been suggested, after performing an empirical study, that keystroke based emotion recognition may be useful in identifying certain emotional states, e.g. anger or sadness. Whereas other emotions such as surprise may be better recognized by facial expression analysis. There are also affective states such as happiness, which are well recognized using both approaches, and disgust which requires another source of information to be well recognized.

The authors of [4] constructed a biometric mouse able to measure the temperature, humidity, skin conductance, touch intensity and heart rate. These physiological parameters were recorded together with standard mouse data such as speed and acceleration of mouse pointer's movements, amplitude of hand tremble, scroll wheel use, right- and left-click frequency and idle time. Before starting to measure all mentioned parameters a user had to fill out a questionnaire about his mood, productivity, stress, self-control, happiness, anger, fear, sadness, surprise, anxiety. The dependency between the self-reported states and the measures parameters was analyzed. The system was used to measure the level of stress and then give some recommendations on how to reduce it.

II. Summary

The presented review leads to some important conclusions. First of all the importance of data collection process has been shown. The samples of a user should be gathered at different times to avoid the situation of getting used to some tasks in the case of recording parameters via specially designed applications. It is also important because of the fact that this type of data is not stable along time. It should be also noted that intentionally induced emotions may cause different reactions than natural circumstances do. Therefore gathering data during usual computer use seems to be better, especially that keyboard and mouse enable us to do it in this way, in contrast to other sources of data applied in affective computing. However it becomes more complicated to label samples collected during normal computer activities.

Another observation is that the way emotions influence the keystroke and mouse parameters often depends on a person. One would type faster in case of stress, another one may pause in such situations. Therefore it is rather recommended to build an individual classifier for every user. This in turns makes the system creator collect larger amounts of training data per person.

As gathering data from keyboard and mouse does not disturb users, it is possible to continuously enlarge training data sets. Therefore a system should be able to update its decision rules in order to improve its performance.

It also turns out that different emotions affect different keystroke and mouse movement parameters. Moreover some of the emotional states are not detected at all in some applications. It means that keystroke dynamics and mouse movements may be indicators of some emotions, but should not be treated as unquestionable. Quite often they should be accompanied by measurements from other sources, e.g. facial expression, voice, natural language etc.

There is still a vast research area to be explored and many questions to be answered. One of possible problems is identifying keystroke or mouse movement features appropriate for the recognition of specified emotional states in given types of applications.

References

[1] Maat L, Pantic M (2007), Gaze-X: Adaptive, Affective, Multimodal Interface for Single-User Office Scenarios, Human Computing, LNAI 4451, Springer-Verlag: 251-271

[2] Szwoch M (2013), FEEDB: a Multimodal Database of Facial Expressions and Emotions, Proceedings of the 6th International Conference on Human System Interaction, Gdańsk

[3] Schuller B, Lang M, Rigoll G (2002), Multimodal emotion recognition in audiovisual communication, Proceedings of IEEE International Conference on Multimedia and Expo, ICME 2002, Lausanne
[4] Kaklauskas A, Zavadskas EK, Seniut M et al (2011), Web-based Biometric Computer Mouse Advisory System to Analyze a User's Emotions and Work Productivity, Engineering Applications of Artificial Intelligence 24: 928-945
[5] Zimmermann P, Gomez P, Danuser B, Schar S (2006), Extending usability: putting affect into the user-experience, Proceedings of the 4th Nordic Conf. on Human-Computer Interaction, Oslo, pp 27-32
[6] Epp C, Lippold M, Mandryk RL (2011), Identifying emotional states using keystroke dynamics, Proceedings of Conf. on Human Factors in Computing Systems, Vancouver, pp 715-724
[7] Lee H, Choi YS, Lee S, Park IP (2012), Towards Unobtrusive Emotion Recognition for Affective Social Communication, Proceedings of the 9th IEEE Consumer Cmmunications and Networking Conference, pp 260-264
[8] Zeng Z, Pantic M, Roisman GI et al. (2009), A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, IEEE Trans. on Pattern Analysis and Machine Intelligence 31(1): 39-58
[9] Calvo RA, D'Mello S (2010), Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications, IEEE Trans. on Affective Computing 1(1): 18-37
[10] Tsoulouhas G, Georgiou D, Karakos A (2011), Detection of Learner's Affective State Based on Mouse Movements, Journal of Computing 3(11): 9-18
[11] Wróbel MR (2013), Emotions in the software development process, Proceedings of the 6th International Conference on Human System Interaction, Gdańsk
[12] Monrose F, Rubin AD (2000), Keystroke dynamics as a biometric for authentication, Future Generation Computing Systems 16: 351-359
[13] Bergadano F, Gunetti G, Picardi C (2002), User authentication through keystroke dynamics, ACM Trans. on Information and System Security 5(4): 367–397
[14] Gunetti G, Picardi C (2005), Keystroke analysis of free text, ACM Trans. on Information and System Security 8(3): 312–347
[15] Dowland PS, Furnell SM (2004), A Long-term Trial of Keystroke Profiling using Digraph, Trigraph and Keyword Latencies, Security and Protection in Information Processing Systems, IFIP 18th WorldComputer Congress, TC11 19th International Information Security Conference, Toulouse
[16] Yampolskiy RV, Govindaraju V (2008), Behavioural biometrics: a survey and classification,Int. Journal of Biometrics 1(1): 81-113
[17] Pusara M, Brodley CE (2004), User re-authentication via mouse movements, Proceedings of the 2004 ACM Workshop on Visualisation and Data Mining for Computer Security, ACM Press, Washington
[18] Ahmed AAE, Traore I (2005), Detecting Computer Intrusions Using Behavioral Biometrics, Third Annual Conference on Privacy, Security and Trust, St. Andrews, New Brunswick, Canada
[19] Bradley M, Lang P (1994), Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry, 25(1):49-59
[20] Likert R (1932), A technique for the measurement of attitudes, Archives of Psychology, 22(140): 1-55
[21] Althothali A (2011), Modeling user affect using interaction events (thesis), University of Waterloo, Canada
[22] Vizer LM, Zhou L, Sears A (2009), Automated stress detection using keystroke and linguistic features, Int. Journal of Human-Computer Studies 67: 870-886
[23] Sottilare RA, Proctor M (2012), Passively Classifying Student Mood and Performance within Intelligent Tutors, Educational Technology & Society 15 (2): 101-114
[24] Khanna P, Sasikumar M (2010), Recognising Emotions from Keyboard Stroke Pattern, Int. Journal of Computer Applications 11(9)
[25] Lv HR, Lin ZL, Yin WJ, Dong J (2008), Emotion recognition based on pressure sensor keyboards, Proceedings of IEEE International Conference on Multimedia and Expo, ICME 2008, Hannover
[26] Maehr W (2005), eMotion - Estimation of the User's Emotional State by Mouse Motions, Diploma thesis for Fachhochschule Vorarlberg, Dornbirn, Austria
[27] Thsihrintzis GA, Virvou M, Alepis E, Stathopoulou IO (2008), Towards improving visual-facial emotion recognition through use of complementary keyboard-stroke pattern information, Proceedings of the Fifth Int. Conf. on Information Technology: New Generations, pp 32-37