Internet Research

Automatic recognition of males and females among web browser users based on behavioural patterns of peripherals usage

Agata Kolakowska Agnieszka Landowska Pawel Jarmolkowicz Michal Jarmolkowicz Krzysztof Sobota

Automatic recognition of males and females among web browser users based on behavioural patterns of peripherals usage

Agata Kolakowska and Agnieszka Landowska
Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Gdansk, Poland, and
Pawel Jarmolkowicz, Michal Jarmolkowicz and Krzysztof Sobota
Harimata Sp. z o.o., Kraków, Poland

Received 7 April 2015
Revised 18 April 2015
8 August 2015
17 October 2015
30 October 2015
Accepted 14 November 2015

Abstract

Purpose – The purpose of this paper is to answer the question whether it is possible to recognise the gender of a web browser user on the basis of keystroke dynamics and mouse movements.
Design/methodology/approach – An experiment was organised in order to track mouse and keyboard usage using a special web browser plug-in. After collecting the data, a number of parameters describing the users' keystrokes, mouse movements and clicks were calculated for each data sample. Then several machine learning methods were used to verify the stated research question.
Findings – The experiment showed that it is possible to recognise males and females on the basis of behavioural characteristics with an accuracy exceeding 70 per cent. The best results were obtained while using Bayesian networks.
Research limitations/implications – The first limitation of the study was the restricted contextual information, i.e. neither the type of web page browsed nor the user activity was taken into account. Another is the narrow scope of the respondent group. Future work should focus on gathering data from more users covering a wider age range and should consider the context.
Practical implications – Automatic gender recognition could be used in profiling a user to create personalised websites or as an additional feature in automatic identification for security reasons. It might be also considered as a confirmation of declared gender in web-based surveys.
Social implications – As not all users perceive personalised ads and websites as beneficial, this application requires the analysis of a user perspective to provide value to the consumer without privacy violation.
Originality/value – Behavioural characteristics, such as mouse movements and keystroke dynamics, have already been used for user authentication and emotion recognition, but applying these data to gender recognition is an original idea.

Keywords Gender recognition, Keystroke dynamics, User modelling, Behavioural patterns, Mouse movements
Paper type Research paper

1. Introduction

User modelling, within the domain of human-computer interaction, focuses on the description of user characteristics in the form of a user model. A user model can be defined as a set of individual user characteristics that are described with attributes or

This research was partly accomplished under the EU EFS Innovation Transfer project (UDA-POKL.08.02.01-12-021/12-00). The authors also thank their colleagues from Emotions in Human-Computer Interaction Research Group at Gdańsk University of Technology (emorg.eu) for their valuable comments on this study.

Internet Research
Vol. 26 No. 5, 2016
pp. 1093-1111

1066-2243
DOI 10.1108/IntR-04-2015-0100

more complex structures. The main purpose of building the model is to be able to adapt and customise system functionality, appearance or behaviour for a particular user profile. User models use diverse characteristics, depending on their application domain and system-specific features (Godoy *et al.*, 2010). The models can be formed by user declaration or automatic recognition. Systems sometimes use the automatic recognition of selected features as a confirmation or replacement of a user declaration. Automatic recognition is usually based on characteristics that are observable in human-computer interactions, including using a keyboard and mouse.

This paper concerns automatic male and female recognition based on behavioural patterns of peripherals usage. Behavioural patterns observable in the use of mouse and keyboard were previously used in person identification (Pusara and Brodley, 2004; Gunetti and Picardi, 2005; Clarke and Furnell, 2006) as well as in emotion recognition (Vizer *et al.*, 2009; Epp *et al.*, 2011; Khanna and Sasikumar, 2010); however, to the best knowledge of the authors of this paper, they have yet to be used in gender recognition. In this paper, the words gender and sex are used interchangeably and refer to male and female distinction; however, the authors are aware of the fact that there could be more to gender interpretation.

The main research hypothesis of this study is as follows:

H1. It is possible to recognise the gender of the user of a web browser by analysing the usage of peripherals.

In order to experimentally verify the hypothesis, data from mouse and keyboard activity were collated with gender labels. This experiment also tried to identify the features which correlate best with gender, as well as the best machine learning methods to solve a given task.

The paper is organised as follows: Section 2 provides a concise summary of the related work; Section 3 describes the research objective, experiment design and methodology; Section 4 contains a preliminary data analysis and the gender recognition results; Section 5 provides a discussion on the proposed solution and Section 6 presents the final conclusions.

2. Related work

The possibility of automatic gender recognition has been of great interest for years. Comprehensive studies on this topic have been presented by Khan *et al.* (2013) or Ng *et al.* (2012), who suggest several areas of possible application, i.e. human-computer interaction, surveillance systems, content-based indexing and searching, biometrics, collecting demographic data, targeted advertising. A great many computer-vision methods have been applied to solve this task. Most of them are based on the analysis of face (Baluja and Rowley, 2007; Shen *et al.*, 2009; Lee *et al.*, 2010; Wu *et al.*, 2011; Shan, 2012), face and hair (Lian and Lu, 2009), or only eyes and eyebrows (Alrashed and Berbar, 2013). Some of them have achieved high accuracies of 99 per cent. There are also solutions based on fingerprints (Gnanasivam and Muttan, 2012; Gornale and Kruthi, 2014), especially important in forensics. Another interesting approach is the analysis of motion sequences presenting gait (Makihara *et al.*, 2010; Hu *et al.*, 2011) or less popular, but also possible, static images of the body (Bourdev *et al.*, 2011). All these methods require a special device, e.g. camera, to obtain source data and their disadvantage is that their performance strongly depends on the orientation, illumination, image background, occlusion, etc.

The proposed method is based on the idea of extracting the information on gender from behavioural data coming from keyboard and mouse. Studies that are mostly related to this research fall into the following categories: research on features describing the usage of peripherals (keyboard and mouse) and studies on how those features can be used in different biometric applications.

Biometric methods, usually used to ensure system security, are based on physical or behavioural features. Physical parameters, such as fingerprints, face, iris, etc., seem to be more appropriate in recognising users due to their stability over time. However, the fact that special hardware is often required to record them is a significant disadvantage. Behavioural characteristics (e.g. voice, handwritten signatures, keystroke dynamics, mouse movements), on the other hand, may be recorded in an unobtrusive way (Yampolskiy and Govindaraju, 2008). Moreover, for some of them it is possible to carry out analysis without disturbing users as they perform their usual tasks. In this sense, standard input devices are especially worth considering. However, the values returned by keyboard or mouse may be easily changed, because they depend on many factors, e.g. the type of hardware and software used, type of user activity or even the user's emotional state.

Behavioural features have been already exploited in different areas. First of all, keystroke dynamics may be used in user authentication and identification (Shanmugapriya and Padmavathi, 2009). No matter if the device to be protected is a computer or a mobile phone, user authentication may be conducted either once during login or continuously upon entering a system. There are solutions to protect mobile phones from intruders while entering PIN codes (BehavioSec, 2012b) and other research investigating the possibility of authentication while entering phone numbers or typing text messages (Clarke and Furnell, 2006). The problem of identifying computer users on the basis of fixed text or any free text has been explored at length by Gunetti and Picardi (2005). Mouse usage may be also treated as an indicator of user characteristics. Pusara and Brodley (2004) authenticated users on the base of parameters obtained for mouse movements, clicks and motion across the menu and toolbar areas using Internet Explorer, whereas Zheng *et al.* (2011) propose a method much more independent from the running environment. The latest research incorporates touch screen biometrics to identify users (Frank *et al.*, 2013; Bo *et al.*, 2013). An interesting approach has been described by Fridman *et al.* (2013), who combined four types of biometric characteristics: keystroke dynamics, mouse movements, stylometry defining a user's linguistic style and web browsing measured by the visit frequency of a few popular websites. This data fusion architecture reduces the error rates achieved using any of the mentioned data sources alone.

Another application, which has become popular lately, is emotion recognition. Systems incorporating both keystroke and mouse movement characteristics have been mentioned in numerous papers (Kołakowska, 2013). It turns out that to some degree it is possible to recognise one (Tsoulouhas *et al.*, 2011; Vizer *et al.*, 2009) or more (Epp *et al.*, 2011; Lee *et al.*, 2012; Schuller *et al.*, 2002; Kołakowska, 2015) emotional states or at least decide whether it is positive or negative (Khanna and Sasikumar, 2010). This may be used in many applications such as, for example, intelligent tutoring systems (Sottilare and Proctor, 2012; Tsoulouhas *et al.*, 2011) or adaptive interfaces (Maat and Pantic, 2006).

Both applications mentioned above reveal that methods based on behavioural features, especially mouse movements (Jorgensen and Yu, 2011), usually give higher error rates than other techniques, e.g. fingerprint authentication or emotion recognition based on facial expression. However, they may be treated as an additional source of information and combined with other modalities (e.g. Karnan and Krishnara, 2012).

User keystroke characteristics are usually described by a number of features, which may be of two types – timing and frequency parameters. Timing features are: typing speed, flight time defined as the time lapse between pressing two subsequent keys and the dwell time which is the time between pressing and depressing a key. Moreover, the duration of key sequences may also be taken into account (Gunetti and Picardi, 2005). Usually digraphs and trigraphs, which are two and three-key sequences, respectively, are considered. The timing parameters may be either averaged over all possible keys or calculated for different keys separately, leading to a large quantity of features. The frequency parameters show how often selected keys are used. Special keys are the most interesting, e.g. the backspace might be used more often by some users and delete by others. The timing parameters describe user keystroke dynamics, whereas the frequency ones may indicate user preferences.

Mouse movement characteristics may be divided into those describing the way a mouse is moved and those describing how the mouse buttons are pressed. Mouse movements are usually characterised by mouse speed, acceleration and a number of parameters connected with movement shapes. They may, for example, measure how the mouse route deviates from a straight line, how often its direction is changed, what the user's preferable direction is, etc. Clicking features may measure the time between pressing and depressing a button or between two presses in the case of a double click. The delays between stopping a mouse and pressing or releasing a button in the case of movement and click or drag and drop events, respectively, are also worth noting. New fine-grained angle-based metrics were also proposed by Zheng *et al.* (2011). It is obvious that many of these features strongly depend on the application used.

The calculated features are usually averaged over sessions containing a series of events. In the case of mouse movements, raw data may also be segmented into strokes giving a feature vector for each stroke.

Although there are no studies on recognising gender from keystroke and mouse characteristics, some interesting conclusions drawn from other research studies may be found. Dora *et al.* (2013), in their paper focusing mainly on user authentication, continuous verification and identification based on keystroke dynamics, present some demographic analysis as well. Their results, comparing the system performance for males and females, indicate that men and women type differently. A broad analysis of different factors on the quality of user authentication based on keystroke dynamics, has been also presented by Killourhy (2012). The results of that experiment show that only typing styles have a significant effect on miss rates while authenticating users, whereas age, gender and the dominant hand do not. The authors do not reject such an influence, but suggest that it might be less significant than in the case of typing styles and it requires further experiments. Such studies reassure the authors of this paper that the idea of recognising users on the basis of behavioural data from input devices is worth investigating.

3. Study design and methodology

3.1 Research objective

This paper aims to verify the idea that keystroke and mouse movement parameters may not only indicate users' identity or emotional state, but also their gender. The authors do not know of any other research exploring such an application. The stated research hypothesis is as follows:

H1. It is possible to recognise the gender of the user of a web browser by analysing the usage of peripherals.

In order to verify this idea, an experiment had to be conducted. The first stage would involve the gathering of data from users of web browsers. Then a set of parameters characterising keystroke dynamics and mouse movements had to be extracted from raw data. Finally machine learning methods would be applied to train gender classifiers and estimate their accuracy.

If the stated hypothesis is confirmed by the experiment results, the authors also intend to identify among all defined mouse and keyboard parameters the most suitable ones to infer gender.

3.2 Data acquisition

In this study, the independent variable is gender, while the dependent variables are the ones that describe behavioural patterns in mouse and keyboard usage. The study was divided into the phases of data acquisition, preprocessing and analysis.

For the data acquisition phase a browser plug-in was used to gather the behavioural characteristics of multiple users. The plug-in logged keystrokes and mouse activities during different activities performed via a browser. Two versions of the plug-in were used for Chrome and Opera. The plug-in required intentional installation and during this phase a user was asked to fill in a short metric with age range, device type and gender declaration. For security reasons the plug-in did not record which alphabetic keys were used and which pages were browsed. The gathered event logs were periodically sent to a server.

In total, 42 people (9 females, 33 males) took part in the data acquisition phase during September and October 2013 providing over 4,300 data samples. Table I presents the distribution of samples among group ages, devices and browsers.

All participants belonged to one of three age groups: 15-24, 25-34, 35-44. Half of the samples came from users from the 25-34 group, but most were male. Most female samples (80 per cent) were from the 15-24 age group, whereas only 7.7 per cent of male samples were from the same group. There were no female examples from the 35-44 group. Had there been samples representing both males and females in all groups, it would have been possible to analyse gender recognition for different age ranges, but the presented distribution does not allow for this.

Most samples, both male and female, were gathered via mouse. There was also a distinct number of samples from touchpad, but most of them were male. Two more input devices, i.e. track point and touch screen, constituted less than 6 per cent of all data. The user was free to change the device used during the experiments. If he or she changed the mouse for some reason, this event was not noted in any way.

The participants used one of two possible web browsers. Almost all used Chrome. Only a few samples were collected from Opera. The users were free to choose a website

	Males (%)	Females (%)	Total (%)
Age			
15-24	7.7	80.0	24.1
15-24	59.6	20.0	50.6
25-34	32.7	0.0	25.3
Device			
Mouse	64.3	88.6	69.8
Touchpad	28.8	9.3	24.4
Trackpoint	6.9	0.0	5.3
Touch screen	0.0	2.1	0.5
Browser			
Chrome	99.8	100.0	99.9
Opera	0.2	0.0	0.1

Table I. Data samples distribution

and no information was available on the type of website used. All participants were experienced users of internet browsers. They could record their characteristics any number of times and whenever they wanted to. Thus the number of samples differed among users.

The raw data from the plug-in included the following events: key down, key up, mouse movement, mouse clicks (different types), as well as scrolls. The events were labelled with time, user id, session id and metric values. The raw data required preprocessing and another application was implemented for this process. A decision was made to calculate a series of features that characterise mouse and keyboard usage patterns. The features were calculated for short sessions instead of individual events. The features that describe mouse usage, are provided in Table II. These parameters were calculated on the basis of studies presented in (Maehr, 2005) and (BehavioSec, 2012a). Table III presents features describing keyboard usage. The choice of features was justified by literature analysis, including features that are used in human identification or emotion recognition.

3.3 Data exploration methodology

The applied research methodology covers data analysis and preprocessing and then training and testing the gender recognition system.

The first step after the feature extraction phase is the statistical analysis of the data. The distribution of samples among males and females, different input devices and various web browsers had to be analysed in order to identify possible unique examples. Moreover, each feature was assessed taking into account the descriptive statistics, e.g. min and max values, mean value, standard deviation, median. One of the important parameters is scarcity defined as the percentage of null values of an attribute among all vectors. Depending on these values a decision on how to deal with missing values should be made.

Table II. Features describing mouse usage

Feature	Description
Acceleration	Acceleration averaged over all movement segments
Deceleration	Deceleration averaged over all movement segments
DirectDistance	Distance between the first and the last event with (x, y) coordinates, e.g. mouse move
MovementAngle	Angle between the horizontal line and the movement direction
MovementEfficiency	The covered distance of a move to the direct distance between the starting and the ending points
MovementSpeed	Speed averaged over movement segments
MovementTargeting	Standard deviation of a movement track from the ideal (straight) movement
MovementUniformity	Measure of speed changes within a movement divided into segments, calculated as standard deviation from the average speed
OvershotNumber	The number of situations when The mouse movement projected onto a line connecting the starting and the ending point moves away from the ending point
OvershotLength	The maximum length of an overshot defined before
Skewness	The total length of a path to the length of its part placed on the left side of a straight line connecting the starting and the ending points
Velocity	The total distance of a move divided by the movement time
MoveAndClickDelay	Time between stopping a mouse and pressing a button
ClickDuration	Time between mousedown and mouseup events
ClickFrequency	Number of clicks in a time unit
ClickPrecision	Distance covered during the click event (other than drag and drop)

After the preprocessing, phase machine learning methods would be applied to train classifiers in a supervised manner and a cross-validation procedure to estimate their accuracies. Two series of tests, based on two different data splitting approaches, were planned in order to evaluate the accuracy of gender recognition. The first tests would apply the standard stratified ten-fold cross-validation. However, the authors are aware of the fact that the results obtained in this way might be optimistically biased, because different samples of a particular user could appear both in the training and testing set. The accuracy evaluated in this way would be the accuracy of a classifier recognising the gender of the 42 participants of the experiment. The recognition accuracy evaluated on the basis of samples coming from males and females other than the 42 experiment participants might be lower. This problem has been also addressed by Baluja and Rowley (2007) or Wu *et al.* (2011), who recognised gender on the basis of facial images. In order to avoid such simplification and obtain a more realistic evaluation of the classifier generalisation ability, another series of tests was planned. In these tests the accuracy would be also estimated in a *k*-fold cross-validation, but in each fold all the samples from one user would be left for testing and the samples of other users for training. The disadvantage of this approach is that the class distributions are not the same in subsequent folds.

The primary parameter used to evaluate the recognition quality is accuracy defined as:

$$\text{accuracy} = \frac{t_{\text{male}} + t_{\text{female}}}{t_{\text{male}} + t_{\text{female}} + f_{\text{male}} + f_{\text{female}}}$$

where t_{male} is the number of correctly recognised male examples, t_{female} – the number of correctly recognised female examples, f_{male} – the number of female examples recognised as male ones and f_{female} – the number of male examples recognised as female ones.

However, intuitive, the accuracy may not be adequate to evaluate and compare different models. Suppose there are 100 samples belonging to two classes and 80 of them belong to class A. If we assigned all samples to class A it would give an accuracy of 80 per cent, which is obviously misleading. So to evaluate the accuracy in each class, the recall parameter, which denotes the probability of correct prediction in a particular class, should be also taken into account:

$$\text{recall}_{\text{female}} = \frac{t_{\text{female}}}{t_{\text{female}} + f_{\text{male}}}$$

$$\text{recall}_{\text{male}} = \frac{t_{\text{male}}}{t_{\text{male}} + f_{\text{female}}}$$

Feature	Description
TypingSpeed[a]	Average time between two subsequent key presses
DwellTime[a]	Time between pressing and depressing a key
FlightTime[a]	Time between pressing two subsequent keys
Backspace %	Number of backspace and delete keys to the total number of keys
SpecialCharPercentage	Number of special character keys to the total number of keys

Note: [a]The following parameters were calculated for every session: average, minimum, maximum, standard deviation, variance, mode, range

Table III. Features describing keyboard usage

Another parameter is precision denoting the probability that a particular prediction (male or female) is correct:

$$\text{precision}_{\text{male}} = \frac{t_{\text{male}}}{t_{\text{male}} + f_{\text{male}}}$$

$$\text{precision}_{\text{female}} = \frac{t_{\text{female}}}{t_{\text{female}} + f_{\text{female}}}$$

Finally the *F*-measure, which is the harmonic mean of precision and recall, may be obtained:

$$F = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

To evaluate how good the accuracy estimation is, a confidence interval should be calculated (Mitchell, 1997). With *N* per cent probability, the true accuracy lies in the following interval:

$$\text{accuracy} \pm z_N \sqrt{\text{accuracy}(1 - \text{accuracy})/n}$$

where n is the number of testing samples; the constant z_N for the 95 per cent confidence interval is 1.96. The next section presents the results of the experiments performed according to the described methodology.

4. Data analysis and results

4.1 Preliminary data analysis

The data represented by feature vectors were carefully analysed to evaluate the data quality and exclude possible anomalies. Preliminary observations on vector distribution led to the following remarks:

- the data were not evenly distributed among the sexes with 906 vectors from female respondents (23 per cent) and 3,087 vectors from male respondents (77 per cent);
- almost all vectors came from the Chrome browser plug-in (only five vectors were obtained from Opera);
- there were four tracking devices used: mouse (69.8 per cent), touchpad (24.3 per cent), trackpoint (5.3 per cent) and touchscreen (less than 0.5 per cent);
- age groups were not evenly represented with 962 vectors from age range 15-24 (24.1 per cent), 2,021 vectors from age group 25-34 (50.6 per cent) and 1,010 vectors from 35-44 (25.3 per cent); and
- there was an uneven distribution of the sexes over age groups.

The observations led to several assumptions that influenced further investigation and recognition processes:

- few data vectors from the Opera plug-in were excluded from further analysis although no difference was expected;
- the three less represented tracking devices (touchpad, trackpoint and tablet) were excluded from the recognition process; and
- the age group attribute was excluded from the recognition process.

As a result of the preprocessing phase, 2,789 vectors from 26 users (4 females, 22 males), containing 803 female and 1,986 male samples, were prepared for the analysis.

The second step of the preliminary data analysis was the evaluation of the numeric attributes, especially concentrating on their scarcity. Some of the attributes had a high scarcity due to the fact that quite often only a mouse was used during one session, leaving the keyboard-related attributes missing. The attributes also differed significantly in range. To address the high range discrepancy among attributes, normalisation was considered for all of them. The means and standard deviations of the attributes must be taken with precaution, as not all distributions were close to normal. The calculated descriptive statistics of selected features are provided in Table IV.

Attribute distribution analysis led to some assumptions concerning further analysis and the recognition process. The features describing typing patterns are scarce, due to the fact that about a half of browser sessions included mouse movements only. Therefore in further analysis three different recognition tracks were performed: for typing attributes only, mouse movement attributes only, combined vectors of mouse and keyboard attributes (with missing values). Selected attributes describing mouse movements were excluded from further analysis due to significant scarcity. As a result of the preprocessing, vectors containing 32 features were prepared for the recognition phase (12 features for mouse movements and 20 for keystrokes).

4.2 Results in recognising males and females

The training and testing were performed twice according to different data splitting approaches, as mentioned in Section 3.3. The first testing procedure was a ten-fold stratified cross-validation. The second was also cross-validation, but in each of its iterations samples of one user were used for testing and all others for training. The data from users who provided the smallest amounts of samples (one to seven feature vectors), were combined into one subset. In this way 21 folds of cross-validation were performed. The number of samples in the subsets differed depending on the amount of data from each user. Each sample was used as a testing vector once.

Attribute	Min	Max	Mean	SD	Median	Scarcity (%)
Acceleration	1.0 E−10	22.7	0.26	0.64	0.12	0
Deceleration	0	2.46	0.09	0.14	0.06	0
DirectDistance	0	137,317	1,351.58	3,900.96	556.94	0
MovementEfficiency	1	70,721	90.99	14,350.48	7.66	38
MovementSpeed	~0	18.16	0.96	0.71	0.82	0
OvershotNumber	1	36,105	1,213.21	3,261.36	280.00	24
OvershotLength	~0	46,408	879.58	1,905.54	455.81	24
Skewness	0	1,674.59	48.82	44.41	46.61	0
Velocity	~0	5.6	0.21	0.38	0.12	0
ClickDuration	1	85,824.00	390.00	2,051.14	115.00	25
ClickFrequency	0	237,600	3,705.65	14,211.71	25.91	26
ClickPrecision	0	15,924.06	68.19	553.76	0	25
Typing speed	~0	23.15	0.41	1.71	0.04	69
DwellTime	~0	97.51	0.27	2.71	0.007	55
FlightTime	~0	12.41	0.06	0.43	0.02	76
Backspace %	0	0.75	0.03	0.08	0	61
SpecialCharPercentage	0	1	0.48	0.38	0.44	61

Table IV. Descriptive statistics of selected features

The data sets were reduced by under-sampling the majority class and preserving the minority one unchanged. This was due to the fact that some machine learning methods are sensitive to an uneven distribution of classes. Moreover, some classifier performance indices, e.g. precision, are also sensitive to class distribution imbalance (Fawcett, 2004). As there was no reason for assuming the dominance of any class, the authors decided to keep the number of samples of males and females equal.

A number of methods with different parameter settings were tested using WEKA software (Hall *et al.*, 2009). The following methods were taken into account: Bayesian networks, neural networks, decision trees and two methods of combining classifiers (AdaBoost, rotation forest). The methods were tested for different parameter settings, i.e. structure learning method and score metric in the case of Bayesian networks, number of layers and neurons in neural networks, attribute selection criterion and pruning algorithm for decision trees. The experiments were performed for mouse movement features, keystroke features and both of them. Principal component analysis was also considered as a feature extraction procedure, but as no result enhancement was observed while using it, the results are not provided. Tables V and VI contain the results of the two testing series obtained for data represented by both mouse and keystroke parameters when the total number of features was 32. The results obtained according to the first data splitting approach are quite satisfying. 80.80 per cent accuracy was obtained while applying rotation forest and 79.12 per cent for the Bayesian network. As presumed, the accuracy evaluated during the second test series was lower. In this more realistic situation, the use of Bayesian networks leads to the best results, with 73.47 per cent samples correctly classified. Rotation forests appear to be slightly worse giving a 70.61 per cent accuracy rate. The results obtained by the other three methods are significantly worse.

Tables VII and VIII present the results achieved for data represented by mouse attributes only. It can be seen that neglecting keyboard parameters reduces the accuracy. In this case the first test series gives a 74.12 per cent level of accuracy. The best result obtained for the more realistic data split is 72.17 per cent. Bayesian networks, again, turn out to be the most appropriate. A series of experiments exclusively for keyboard features was also performed, but the results obtained in this

	Bayesian network	Rotation forest	Decision tree	AdaBoost	Neural network
Accuracy (%)	79.12	80.80	73.89	77.40	75.80
95% confidence interval (%)	77.13-81.11	78.87-82.73	71.74-76.04	75.35-79.44	73.70-77.89
Recall (%)					
Male	76.80	80.20	72.02	74.47	72.31
Female	81.44	81.40	75.76	80.32	79.29
Avg.	79.12	80.80	73.89	77.40	75.80
Precision (%)					
Male	80.54	81.18	74.85	79.10	77.73
Female	77.83	80.43	73.04	75.88	74.12
Avg.	79.18	80.81	73.94	77.49	75.93
F-measure (%)					
Male	78.62	80.69	73.41	76.72	74.92
Female	79.59	80.91	74.38	78.04	76.62
Avg.	79.11	80.80	73.89	77.38	75.77

Table V. Results obtained using both mouse and keyboard features during the first test series (ten-fold cross-validation)

	Bayesian network	Rotation forest	Decision tree	AdaBoost	Neural network
Accuracy (%)	73.47	70.61	66.31	63.57	63.64
95% confidence interval (%)	71.32-75.63	68.38-72.84	64.00-68.63	61.22-65.93	61.28-65.99
Recall (%)					
Male	75.34	72.10	68.74	67.87	64.01
Female	71.61	69.12	63.89	59.28	63.26
Avg.	73.47	70.61	66.31	63.57	63.64
Precision (%)					
Male	72.63	70.01	65.56	62.50	63.54
Female	74.39	71.25	67.15	64.85	63.74
Avg.	73.51	70.63	66.35	63.68	63.64
F-measure (%)					
Male	73.96	71.04	67.11	65.07	63.77
Female	72.97	70.16	65.48	61.94	63.50
Avg.	73.47	70.60	66.29	63.51	63.64

Table VI. Results obtained using both mouse and keyboard features during the second test series (one user left for testing in each iteration)

	Bayesian network	Rotation forest	Decision tree	AdaBoost	Neural network
Accuracy (%)	74.12	73.97	72.35	72.93	72.02
95% confidence interval (%)	71.98-76.26	71.83-76.12	70.17-74.54	70.76-75.11	69.83-74.22
Recall (%)					
Male	71.32	72.40	72.02	70.90	72.10
Female	76.92	75.55	72.69	74.97	71.94
Avg.	74.12	73.97	72.35	72.93	72.02
Precision (%)					
Male	75.55	74.76	72.53	73.92	72.01
Female	72.85	73.25	72.31	72.04	72.05
Avg.	74.20	74.01	72.42	72.98	72.03
F-measure (%)					
Male	73.37	73.56	72.28	72.38	72.06
Female	74.83	74.38	72.50	73.48	71.99
Avg.	74.10	73.97	72.39	72.93	72.03

Table VII. Results obtained using mouse features only during the first test series (ten-fold cross-validation)

way were significantly worse. One of the reasons might be the significant number of missing values in the case of keystroke characteristics due to the fact that some users did not use the keyboard at all. All applied algorithms, implemented in WEKA, cope with missing values, but in different ways. Bayesian networks, for example, fill the missing values with the means estimated on the basis of the training data. Neural networks ignore them and replace them with zeros. Decision trees (C4.5 in this case) and rotation forests, which combine the same C4.5 decision trees, modify the value of the attribute selection criterion on the basis of the proportion of missing values while in the testing phase they incorporate weights for all the possible leaves reached by a testing example. AdaBoost treats the missing values as a separate value. These approaches allow a classifier to be trained and tested, but if the number of missing values is

Table VIII. Results obtained using mouse features only during the second test series (one user left for testing in each iteration)

	Bayesian network	Rotation forest	Decision tree	AdaBoost	Neural network
Accuracy (%)	72.17	69.80	65.75	62.39	61.52
95% confidence interval (%)	69.97-74.36	67.56-72.05	63.43-68.07	60.02-64.76	59.14-63.90
Recall (%)					
Male	72.73	70.49	67.62	65.50	60.52
Female	71.67	69.12	63.89	59.28	62.52
Avg.	72.17	69.80	65.75	62.39	61.52
Precision (%)					
Male	71.92	69.53	65.19	61.66	61.75
Female	72.42	70.08	66.36	63.21	61.29
Avg.	72.17	69.80	65.78	62.44	61.52
F-measure (%)					
Male	72.32	70.01	66.38	63.53	61.13
Female	72.01	69.59	65.10	61.18	61.90
Avg.	72.17	69.80	65.74	62.35	61.52

relatively high it may strongly influence the results. Therefore, the accuracy increase while using all features is usually rather low (1-2 per cent). It appears more significant only for Bayesian networks and rotation forests in the case of the first test series.

The 95 per cent confidence intervals were calculated to evaluate the precision of the obtained accuracy estimation. As may be observed from Tables V-VIII, in all cases the width of the confidence interval varies from 4 to nearly 5 per cent. This means that with a 0.95 probability the true accuracy does not differ from the estimated one more than 2-2.5 per cent, which may be treated as a precise estimation.

Another observation worth noting is that the recognition accuracy strongly varies among users, which could be evaluated on the basis of the second test series. There are users whose samples are recognised with a 90 per cent accuracy and those, for whom the results are no better than guesswork. This is a typical observation while working with behavioural data, made also during the authentication task. Some users' mouse movements and keystrokes are stable over time, whereas others are significantly influenced by different factors such as, for example, the type of web page browsed, type of device used, emotional state, etc.

Apart from classification accuracy, some of the applied knowledge representations provide additional, valuable information. Decision trees, for example, may be treated as a model with an embedded feature selection mechanism. Each node of a tree contains a feature, which is optimal in the sense of feature selection criterion used during the tree induction procedure. The closer from the root node, the higher the discriminative power of the feature. Precise analysis of the trees constructed using the Gini index or gain ratio as feature selection criteria show that the best two features are Deceleration and MovementSpeed, both coming from mouse. The next several features used on high levels of the trees are: Acceleration, Velocity, MovementEfficiency and Direct Distance. After constructing a tree only on the base of keystroke parameters it turns out that the optimal features are: SpecialCharacterPercentage, MinTypingSpeed, AvgDwellTime and TypingSpeedVariance.

These observations provoked a more precise analysis of the attributes' discriminative power. In order to compare the different characteristic features of males and females, an analysis of variance (ANOVA) was performed, comparing intragroup variability with intergroup variability in order to confirm the existence of differences between groups. The one-way ANOVA analysis was performed for every feature and the results (*F*-Snedecor test) are provided in Table IX (significant results only). The *F*-Snedecor test did not confirm the significance of other features, i.e. the intragroup variability was higher than among groups for other features.

5. Discussion

The experiments confirmed the idea that mouse movements and keystroke characteristics may provide useful information in gender recognition tasks. The best result was achieved taking into account both types of features together and applying Bayesian networks as a classification method. This configuration led to an overall accuracy rate of 73.47 per cent. The results obtained by rotation forests were only slightly worse, but the results from other methods were unsatisfactory. Moreover, precise analysis of the constructed decision trees showed the best features in this application. Although the recognition rate obtained so far may not be high enough in some applications, the authors believe the method has some potential. Thus, the research will be continued. Various classifiers with different parameter settings have been tested, but there are still some approaches worth exploring, which could probably enhance the results. Special attention will be paid to the possibility of defining new features. Moreover it is possible to incorporate some other available information into feature vectors, e.g. type of website used.

The obtained level of accuracy is quite low if compared, for example, to gender recognition methods based on facial analysis. However, it should be highlighted that the behavioural data analysed in the presented experiment, i.e. keystroke and mouse characteristics, used in other applications also may not appear to be the most accurate. Authentication based on face, fingerprint or iris usually gives lower error rates. Recognising emotions on the basis of face expression or language is also more reliable. The reason for this is the nature of behavioural data, which are unstable along time and may be influenced by lots of factors. Different keyboards and mice, type of software used and activity performed, even user posture or environmental conditions may influence the users' typing rhythm, cursor movements speed, smoothness, etc. Their main advantage, on the other hand, is the possibility of an unobtrusive and continuous data recording process. Moreover they may be treated as an additional source of

Feature	*F*-test parameters df	Threshold ($p < 0.05$)	*F*-Snedecor test results *F*	*p*-value	Significance with $p < 0.05$	Significance with $p < 0.01$
Deceleration	1/420	3.860	23.94	0.000001415	Y	Y
MovementSpeed	1/420	3.860	33.52	0.000000013	Y	Y
OvershotLength	1/340	3.873	7.77	0.005598684	Y	Y
AvgDwellTime	1/420	3.860	11.18	0.000902389	Y	Y
DwellTimeStdDev	1/420	3.860	13.61	0.000254283	Y	Y
DwellTimeVariance	1/420	3.860	9.86	0.001812099	Y	Y
MaxDwellTime	1/420	3.860	5.99	0.014747590	Y	
DwellTimeRange	1/420	3.860	5.83	0.016178506	Y	
FlightTimeStdDev	1/293	3.888	5.30	0.021987402	Y	
SpecialCharPercentage	1/378	3.873	14.05	0.000206058	Y	Y

Table IX. *F*-Snedecor test results on significance of selected features' differences between males and females information on the users' gender. There are a lot of studies reporting a decrease in error rates after combining the results of various methods incorporating different modalities, both in the areas of authentication (Fridman *et al.*, 2013) and emotion recognition (Calvo and D'Mello, 2010). This may also be true for gender recognition and should be verified by particular experiments.

The experiment results allow the hypothesis, that, to some extent, it is possible to recognise gender by analysing the usage of peripherals. However, we acknowledge that our approach to data collection and study analysis are not without reservations. This section reflects the main limitations of the study, i.e. restricted contextual information, the narrow scope of the respondent group and application constraints.

5.1 Limited contextual information

One limitation of the data collection method chosen for the study is that the web browser plug-in collects information on keyboard and mouse usage independently of the webpage explored by a user. Only the information on the type of web browser is given. Therefore, the data from the usage of peripherals do not reflect either the type of a webpage browsed or the kind of actions performed. Moreover, for security and privacy reasons, the key codes are not logged (only special keys, such as Backspace, are logged). The data do not include information on the place where the computer was used and it could be expected that behavioural patterns at a computer desk differ from, for example, the train. With the context of use, we could answer the question whether we are able to separate behavioural features that change with context from the characteristics that are context-independent. Having the data without context information we cannot extract context-dependent behavioural patterns.

5.2 Limitations of the scope of the study group

In total, 42 people took part in the data collection process. The construction of this group was limited as not all possible confounding variables values are represented. The study was performed in Poland, among healthy adult volunteers aged from 21 up to 44, who are experienced users of internet browsers. The main limitation is lack of diversity of race and nationality, although it might be expected that behavioural patterns in mouse and keyboard usage are the same in Poland as in other developed countries. Cultural and temperamental differences may influence the behavioural patterns, therefore the potential to apply the study to the entire human population is limited. There was also no representation of people aged over 44 years so it was impossible to make any assumptions on this age group. Physiology changes with age and so may the behavioural patterns associated with mouse and keyboard usage. Moreover, elderly people often do not use technology and computers on a daily basis and, therefore, their patterns may differ significantly. Some recognisable differences in behavioural patterns among age groups have been observed; however, in this study age groups were not evenly distributed, so the result should be taken with precaution and confirmed by a separate study.

5.3 Theoretical and practical implications

The most important practical implication of the study is that it provides an additional source of information on users' gender, apart from directly asking them to state their gender in web-based forms. In such a context, the obtained results might be interesting for web page developers, portal owners and researchers running web surveys. It opens the possibility of getting the information when it is desired rather than being provided by a user. Forms asking users for personal information such as gender, age or educational status are often perceived as intrusive. People often ignore them or enter incorrect data (Akbulut, 2015). Relying on personal data given by respondents while running any studies might be risky. The problem of the quality and credibility of data coming from social media should be addressed with special methods (Schoen *et al.*, 2013). If information on gender is crucial for the conducted research and there is a risk of intentional misinformation in this matter, one might use automatic recognition as a confirmation method. If there is a discrepancy in the declared and the recognised sex, the decision on the interpretation should depend on the purpose of the study. The scale and background of the inconsistencies might be relevant from the perspective of gender studies.

The presented research can be also directly applied in user modelling. A lot of research effort is put into creating systems or websites adapting to users' needs and preferences (e.g. Maat and Pantic, 2006; Kao and Wu, 2012; Fong *et al.*, 2011; Godoy *et al.*, 2010; Kathuria *et al.*, 2010; Han *et al.*, 2014). Incorporating gender recognition could enrich some solutions. User profiling, however, the most obvious and straightforward, is not the only possible application of the presented recognition method. It may be also used as a step in the user authentication process, especially when combined with other verification methods. In that application area gender recognition perhaps would not be perceived as intrusive, as customers are interested in the high accuracy and recall of the authentication process.

Nowadays with Google intensive personalisation, users are more or less aware of the fact that their preferences are monitored and individual profile is recognised (Pariser, 2011). The problem of privacy concern is especially visible in social networking, which is a vast source of information. Even if users are aware of possible threats and express their concern, they still accept using such websites (Tan *et al.*, 2012). As some users perceive personalised adds and websites as beneficial, while others as threatening or frustrating, this application case requires the analysis of the user perspective, so that the tool could provide value for the consumer. There are number of studies on the best practices and norms in privacy assurance (Nissenbaum, 2004).

The authors are aware of the fact that the result provided carries the possible threat of misuse. The main concern includes privacy violation and use as a discrimination tool. However, the authors believe that it might be used also in ethically sound ways and be worthwhile for the person it recognises.

6. Conclusions

The paper presents the new idea of on-line gender recognition. Its novelty arises from the type of data used to infer about gender. In contrast to frequently used methods based mainly on face recognition, this approach incorporates behavioural characteristics coming from mouse and keyboard, which have been already applied in user authentication and emotion recognition.

The results of the presented research motivate to perform some further investigations. The first idea is to incorporate the knowledge about the type of web page used and thus the type of user activity, which may strongly influence the recorded characteristics. Moreover the same analysis could be made concerning applications other than web browsers. Then it would be also worth experimenting on different age groups covering a wider age range.

The replication of the data collection in different settings, with more contextual information and a broader population sample would constitute a most interesting study that would lead to the exploration of cultural differences as well as the ageing of behavioural patterns. Finally, the revealed behavioural patterns may indeed be interesting for gender studies. To sum up, the various potential applications and the quantity of intriguing ideas to be tested make the results interesting, although ethical concern is advised in any area of application.

References

Akbulut, Y. (2015), "Predictors of inconsistent responding in web surveys", *Internet Research*, Vol. 25 No. 1, pp. 131-147.

Alrashed, H.F. and Berbar, M.A. (2013), "Facial gender recognition using eyes images", *International Journal of Advanced Research in Computer and Communication Engineering*, Vol. 2 No. 6, pp. 2441-2445.

Baluja, S. and Rowley, H.A. (2007), "Boosting sex identification performance", *International Journal of Computer Vision*, Vol. 71 No. 1, pp. 111-119.

BehavioSec (2012a), "Mouse dynamics, behaviometrics, a paradigm shift in computer security", white paper, BehavioSec, Stockholm, available at: www.behaviosec.com/wp-content/uploads/2012/11/ (accessed 13 November 2013).

BehavioSec (2012b), "BehavioMobile, applying the BehavioSec technology for multilayered mobile security", white paper, BehavioSec, Stockholm, available at: www.behaviosec.com/products/mobile-authentication/ (accessed 17 April 2014).

Bo, C., Zhang, L., Li, X.-Y., Huang, Q. and Wang, Y. (2013), "SilentSense: silent user identification via dynamics of touch and movement behavioral biometrics", *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, pp. 187-190.

Bourdev, L., Maji, S. and Malik, J. (2011), "Describing people: a Poselet-based approach to attribute classification", *Proceedings of IEEE International Conference on Computer Vision*, pp. 1543-1550.

Calvo, R.A. and D'Mello, S. (2010), "Affect detection: an interdisciplinary review of models, methods, and their applications", *IEEE Transactions on Affective Computing*, Vol. 1 No. 1, pp. 18-37.

Clarke, N.L. and Furnell, S.M. (2006), "Authenticating mobile phone users using keystroke analysis", *International Journal of Information Security*, Vol. 6 No. 1, pp. 1-14.

Dora, R.A., Schalk, P.D., McCarthy, J.E. and Young, S.A. (2013), "Remote suspect identification and the impact of demographic features on keystroke dynamics", *Proceedings of SPIE 8757, Cyber Sensing, Baltimore, April.*

Epp, C., Lippold, M. and Mandryk, R.L. (2011), "Identifying emotional states using keystroke dynamics", *Proceedings of Conference on Human Factors in Computing Systems, Vancouver*, 2007, pp. 715-724.

Fawcett, T. (2004), *ROC Graphs: Notes and Practical Considerations for Researchers*, Hewlet-Packard Company, Palo Alto, CA.

Fong, A., Zhou, B., Hui, S., Hong, G. and Do, T.A. (2011), "Web content recommender system based on consumer behavior modeling", *IEEE Transactions on Consumer Electronics*, Vol. 57 No. 2, pp. 962-969.

Frank, M., Biedert, R., Ma, E., Martinovic, I. and Song, D. (2013), "Touchalytics: on the applicability of touchscreen input as a behavioral biometric for continuous authentication", *IEEE Transactions on Information Forensics and Security*, Vol. 8 No. 1, pp. 136-148.

Fridman, A., Stolerman, A., Acharya, S., Brennan, P., Juola, P., Greenstadt, R. and Kam, M. (2013), "Decision fusion for multimodal active authentication", *IT Professional*, Vol. 15 No. 4, pp. 29-33.

Gnanasivam, P. and Muttan, S. (2012), "Fingerprint gender classification using wavelet transform and singular value decomposition", *International Journal of Computer Science Issues*, Vol. 9 No. 2, pp. 274-282.

Godoy, D., Schiaffino, S. and Amandi, A. (2010), "Integrating user modeling approaches into a framework for recommender agents", *Internet Research*, Vol. 20 No. 1, pp. 29-54.

Gornale, S.S. and Kruthi, R. (2014), "Fusion of fingerprint and age biometric for gender classification using frequency and texture analysis", *Signal & Image Processing: An International Journal*, Vol. 5 No. 6, pp. 75-85.

Gunetti, G. and Picardi, C. (2005), "Keystroke analysis of free text", *ACM Transactions on Information and System Security*, Vol. 9 No. 3, pp. 312-347.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009), "The WEKA data mining software: an update", *SIGKDD Explorations*, Vol. 11 No. 1, pp. 10-18, available at: www.cs.waikato.ac.nz/ml/weka/

Han, J., Schmidtke, H.R., Xie, X. and Woo, W. (2014), "Adaptive content recommendation for mobile users: ordering recommendations using a hierarchical context model with granularity", *Pervasive and Mobile Computing*, Vol. 13, August, pp. 85-98.

Hu, M., Wang, Y., Zhang, Z. and Zhang, D. (2011), "Gait-based gender classification using mixed conditional random field", *IEEE Transactions on Systems, Man, and Cybernetics, Part B, Cybernetics*, Vol. 41 No. 5, pp. 1429-1439.

Jorgensen, Z. and Yu, T. (2011), "On mouse dynamics as a behavioral biometric for authentication", *Proceedings of the 6th ACM Symposium on Information, Computer and Communication Security*, pp. 476-482.

Kao, S. and Wu, C. (2012), "PIKIPDL: a personalized information and knowledge integration platform for DL service", *Library Hi Tech*, Vol. 30 No. 3, pp. 490-512.

Karnan, M. and Krishnara, N. (2012), "A model to secure mobile devices using keystroke dynamics through soft computing techniques", *International Journal of Soft Computing and Engineering*, Vol. 2 No. 3, pp. 71-75.

Kathuria, A., Jansen, B.J., Hafernik, C. and Spink, A. (2010), "Classifying the user intent of web queries using k-means clustering", *Internet Research*, Vol. 20 No. 5, pp. 563-581.

Khan, S.A., Ahmad, M., Nazir, M. and Riaz, N. (2013), "A comparative analysis of gender classification techniques", *International Journal of Bio-Science and Bio-Technology*, Vol. 5 No. 5, pp. 223-243.

Khanna, P. and Sasikumar, M. (2010), "Recognising emotions from keyboard stroke pattern", *International Journal of Computer Applications*, Vol. 11 No. 9, pp. 1-5.

Killourhy, K.S. (2012), "A scientific understanding of keystroke dynamics", doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Kołakowska, A. (2013), "A review of emotion recognition methods based on keystroke dynamics and mouse movements", *Proceedings of the 6th International Conference on Human System Interaction*, pp. 548-555.

Kołakowska, A. (2015), "Recognizing emotions on the basis of keystroke dynamics", *Proceedings of the 8th International Conference on Human System Interaction*, pp. 291-297.

Lee, H., Choi, Y.S., Lee, S. and Park, I.P. (2012), "Towards unobtrusive emotion recognition for affective social communication", *Proceedings of the 9th IEEE Consumer Communications and Networking Conference*, pp. 260-264.

Lee, P.H., Hung, J.Y. and Hung, Y.P. (2010), "Automatic gender recognition using fusion of facial strips", *Proceedings of the 20th International Conference on Pattern Recognition*, pp. 1140-1143.

Lian, X.C. and Lu, B.L. (2009), "Gender classification by combining facial and hair information", *15th International Conference, ICONIP 2008, Advances in Neuro-Information Processing, Lecture Notes in Computer Science, Part II*, *Vol. 5507*, *Auckland*, pp. 647-654.

Maat, L. and Pantic, M. (2006), "Gaze-X: adaptive affective multimodal interface for single-user office scenarios", *Proceedings of the 8th International Conference on Multimodal Interfaces*, pp. 171-178.

Maehr, W. (2005), "eMotion – estimation of the user's emotional state by mouse motions", diploma thesis, Fachhochschule Vorarlberg, Dornbirn.

Makihara, Y., Mannami, H. and Yagi, Y. (2010), "Gait analysis of gender and age using a largescale multi-view gait database", *Proceedings of the 10th Asian Conference on Computer Vision*, pp. 440-451.

Mitchell, T.M. (1997), *Machine Learning*, McGraw-Hill, Inc, New York, NY.

Ng, C.B., Tay, Y.H. and Goi, B.-M. (2012), "Vision-based human gender recognition: a survey", Computing Research Repository (CoRR), arXiv:1204.1611, Cornell University Library.

Nissenbaum, H. (2004), "Privacy as contextual integrity", *Washington Law Review*, Vol. 79 No. 1, pp. 119-158.

Pariser, E. (2011), "The troubling future of internet search", *The Futurist*, September-2 October, pp. 6-7.

Pusara, M. and Brodley, C.E. (2004), "User re-authentication via mouse movements", *Proceeding of the 2004 ACM Workshop on Visualisation and Data Mining for Computer Security*, pp. 1-8.

Schoen, H., Gayo-Avello, D., Takis Metaxas, P., Mustafaraj, E., Strohmaier, M. and Gloor, P. (2013), "The power of prediction with social media", *Internet Research*, Vol. 23 No. 5, pp. 528-543.

Schuller, B., Lang, M. and Rigoll, G. (2002), "Multimodal emotion recognition in audiovisual communication", *Proceedings of IEEE International Conference on Multimedia and Expo*, *Lausanne*, pp. 745-748.

Shan, C. (2012), "Learning local binary patterns for gender classification on real-world face images", *Pattern Recognition Letters*, Vol. 33 No. 4, pp. 431-437.

Shanmugapriya, D. and Padmavathi, G. (2009), "A survey of biometric keystroke dynamics: approaches, security and challenges", *International Journal of Computer Science and Information Security*, Vol. 5 No. 1, pp. 115-119.

Shen, B.C., Chen, C.S. and Hsu, H.H. (2009), "Fast gender recognition by using a shared-integral-image approach", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei*, pp. 521-524.

Sottilare, R.A. and Proctor, M. (2012), "Passively classifying student mood and performance within intelligent tutors", *Educational Technology & Society*, Vol. 15 No. 2, pp. 101-114.

Tan, X., Qin, L., Kim, Y. and Hsu, J. (2012), "Impact of privacy concern in social networking web sites", *Internet Research*, Vol. 22 No. 2, pp. 211-233.

Tsoulouhas, G., Georgiou, D. and Karakos, A. (2011), "Detection of learner's affective state based on mouse movements", *Journal of Computing*, Vol. 3 No. 11, pp. 9-18.

Vizer, L.M., Zhou, L. and Sears, A. (2009), "Automated stress detection using keystroke and linguistic features", *International Journal of Human-Computer Studies*, Vol. 67 No. 10, pp. 870-886.

Wu, T.X., Lian, X.-C. and Lu, B.-L. (2011), "Multi-view gender classification using symmetry of facial images", *Neural Computing and Applications*, Vol. 21 No. 4, pp. 661-669.

Yampolskiy, R.V. and Govindaraju, V. (2008), "Behavioural biometrics: a survey and classification", *International Journal of Biometrics*, Vol. 1 No. 1, pp. 81-113.

Zheng, N., Paloski, A. and Wang, H. (2011), "An efficient user verification system via mouse movements", *Proceedings of the 18th ACM Conference on Computer and Communications Security*, pp. 139-150.

About the authors

Agata Kolakowska is an Assistant Professor at the Department of Intelligent Interactive Systems, Faculty of Electronics, Telecommunications and Informatics of Gdańsk University of Technology (GUT). She received a PhD Degree in Computer Science from the GUT. Her research focuses on machine learning methods applied in computer recognition systems, taking into account data collection and preprocessing, feature selection and extraction, training and testing. She takes part in several research projects on emotion recognition and behavioural biometrics. She is a Member of Emotions in HCI Research Group. Agata Kolakowska is the corresponding author and can be contacted at: agatakol@eti.pg.gda.pl

Agnieszka Landowska is an Assistant Professor at the Department of Software Engineering at the Gdańsk University of Technology (GUT). She received a PhD Degree in information technology from the GUT. She is a Leader of Emotions in HCI Research Group and her research concentrates on software user experience testing based on effect as well as effective tutoring systems design. She manages projects for effective methods development and supporting autistics therapy. She is an Editor in Chief of scientific journal *EduAction: Electronic Education Magazine*. She is a Member of Association for Advances in Affective Computing and a Board Member of Polish Society for e-learning.

Pawel Jarmolkowicz is a Founder and VP at Harimata company that developed a research-based technology that combines advances in behavioural science and artificial intelligence studies for early assessment of developmental disorders in children. Computer scientist by education (Eng.), entrepreneur and experienced innovator. Co-founded diagnostic startup Hoope, and led large developer teams providing solutions for the largest financial institutions in Poland. Graduate of Singularity University (2014), think-tank to seek technological solutions to the world's greatest challenges.

Michal Jarmolkowicz is currently an Advisory IT Specialist at IBM, specializing in Cryptography, Pattern Discovery and Pattern Recognition. He received his MSc in Safe and Secure IT Systems from the Technical University of Denmark. His Master Thesis (A Grid-aware Intrusion Detection System, IMM-Thesis-2007-109) focused on feature selection and performance of pattern discovery algorithms in network traffic analysis. His Bachelor Thesis (Converting Print Music Notation to Braille Music Notation, Polish State Committee for Scientific Research, 3T11C 009 26) focused on recognition and processing of music notation.

Krzysztof Sobota Graduated from the University of Science and Technology in Kraków, with Bachelor Degree in Applied Computer Science. As a part of his research he currently works on the engine that detects anomalies in movement patterns related to autism with usage of mobile devices. He is interested in software engineering and data analysis.