Graphology based handwritten character analysis for human behaviour identification

Subhankar Ghosh¹, Palaiahnakote Shivakumara²✉, Prasun Roy¹, Umapada Pal¹, Tong Lu³

¹Indian Statistical Institute, Kolkata 700108, India
²Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
³National Key Lab for Novel Software Technology, Nanjing University, 210093 Nanjing, People’s Republic of China

✉ E-mail: hudempsk@yahoo.com

Abstract: Graphology-based handwriting analysis to identify human behavior, irrespective of applications, is interesting. Unlike existing methods that use characters, words and sentences for behavioural analysis with human intervention, we propose an automatic method by analysing a few handwritten English lowercase characters from a to z to identify person behaviours. The proposed method extracts structural features, such as loops, slants, cursive, straight lines, stroke thickness, contour shapes, aspect ratio and other geometrical properties, from different zones of isolated character images to derive the hypothesis based on a dictionary of Graphological rules. The derived hypothesis has the ability to categorise the personal, positive, and negative social aspects of an individual. To evaluate the proposed method, an automatic system is developed which accepts characters from a to z written by different individuals across different genders and age groups. This automatic privacy projected system is available on the website (http://subha.pythonanywhere.com). For quantitative evaluation of the proposed method, several people are requested to use the system to check their characteristics with the system automatic response based on his/her handwriting by choosing to agree or disagree options. The automatic system receives 5300 responses from the users, for which, the proposed method achieves 86.70% accuracy.

1 Introduction

Graphology based behavioural analysis is gaining popularity in the recent years due to widespread applications across diverse fields, such as psychology, education, medicine, criminal detection, marriage guidance, commerce and recruitment [1], etc. In addition, it is also noted that graphology based handwriting reveals inner feelings of persons though such characteristics are invisible from person behaviours [2]. Therefore, traditional methods that use visible facial/biometric features or human actions to identify person behaviours may not be effective, especially when a person pretends artificially. Besides, most conventional methods are application, situation and dataset dependent. Thus, graphology based handwriting analysis is used as an objective tool for studying person behaviours without depending on appearance-based features of persons to make a system independent on fields, data, gender, age of a person, applications, etc. Further, since graphology focuses on individual letters, strokes and part of a character rather than the whole character, word, or document, features will be sensitive to personal behaviours, which help in predicting person behaviours [1]. Several methods have been proposed for predicting person behaviours using graphology based handwriting in the literature [1, 2]; however, such methods expect the human intervention to identify behaviours. Therefore, there is an urgent need for developing a generalised graphology based handwriting analysis method. To study behaviours such as emotions, feelings and person identifications, there are methods [3, 4] which use a signature for identification. However, these methods consider the whole signature for prediction in contrast to graphology which considers a part of a character. Similarly, there are studies on aesthetic analysis [5, 6], which use handwritten characters or image features to predict personal behaviours such as beautiful, non-beautiful, excellent writing and poor writing. However, these methods are limited to few behaviour types. There are methods proposed for personality assessment [7, 8], which assess person interest, attitude, relationship with family, community, etc. These methods study typed texts by users but not handwritten texts. In general, the methods require full-text lines or sentences for personality assessment. Therefore, to predict or identify personal behaviours which may be invisible from eyes, graphology based handwriting analysis becomes essential.

It is noted that graphology based human behaviour analysis does not have a mathematical basis. However, the rules and definition are derived based on the discussions and views of the people of a Graphological Institute in Kolkata (http://mbose-kig.com/, India). Also, some of the rules are obtained based on the experience and psychology of the people, as discussed in the link (https://ipip.org/). It is evident from the literature [7, 8], where we can see several methods are published for studying human behaviour using graphology. For example, when the person is under pressure and in poor condition, it is expected such behaviour re...
interest for feature extraction. As a result, if the method does not get the correct regions of interest, it may not perform well. In addition, the features extracted are sensitive to background and foreground. Fallah and Khotanlou [9] proposed to identify human personality parameters based on handwriting and neural networks. The method extracts writing style to find the personality parameters of a person. In other words, the method finds variations in writing using correlation estimation and feature extraction. Extracted features are fed to a neural network classifier for personality parameters detection. However, it is not clear about the number of parameters and the basis for parameter selection. In addition, the method requires at least one word written by a person for personality parameter detection. Though, the above two methods are related to person behaviour identification, the scopes and the ways they extract features are different from the proposed method.

Topaloğlu and Ekmecki [10] proposed gender detection for handwriting analysis. The method extracts attributes of handwritten characters for identifying gender, male or female. The method works based on the fact that texts written by a female are often neat, visible and legible, and one can expect uniform spacing between words, text lines, etc. While from male writing, it is hard to find the above characteristics. Therefore, the method extracts pressure, border, space, the dimension of baselines, slanting, etc. In total, the method extracts 133 attributes and then uses a decision tree for classification. Though the method studies graphology using handwritten characters, the extracted attributes are limited to two classes. In addition, their scope is gender classification but not personal behaviour identification.

Champa and Kumar [11] proposed artificial neural networks for human behaviour prediction through handwriting analysis. The method extracts baseline with its inclination and writing pressure to identify personal behaviours such as different levels of emotions and confidence. For this purpose, the method uses a character ‘t’. In other words, the method is limited to study the attributes of character ‘t’. However, the scope of the proposed work is to study the attributes of characters to identify different types of personal behaviours.

Champa and Kumar [12] also proposed automated human behaviour prediction through handwriting analysis. The above authors developed a method for identifying personal traits using handwriting character analysis. In this work, the method explores generalised Hough transform for studying the orientation of character ‘y’ is either left, right vertical or extreme right-skewed. Comparing to [11], this work focuses on the attributes of character ‘y’. Therefore, the scope is limited to specific characters but not a generalised method.

Coll et al. [13] proposed a graphological analysis of handwritten text documents for human resource recruitment. In this work, the method focuses on attributes, such as active personality and leadership, which are required for human resource recruitment. The features such as layout configuration, letter size, shape, slant and skew angle of lines, are considered for the above personal trait identification. The method explores projection profile based features, contour-based features, discrete cosine transform and entropy for extracting the above features from handwritten documents. It is noted that the method requires the full document and focuses only on the requirement of human resource recruitment. Therefore, their scope is limited to specific applications.

In summary, the methods [11–13] focus on extracting features such as writing force, pressure as well as shapes of characters. Besides, the scopes of the methods are limited to specific behaviours and particular characters but not generalised behaviours with personality characters as the proposed work. This is because the methods are proposed for particular applications. Further, the features extracted based on writing force and pressure may not be effective compared to shape-based features as one can expect the same from different persons. In addition, pressure and force depend on paper quality, thickness, pen, ink, etc. On the other hand, in contrast to the existing methods [11–13], the proposed system does not target any particular application and is not limited to specific characters. As a result, the way the proposed method extracts features and defines rules based on graphologist is different from the above-mentioned existing methods. The key advantage of the proposed system is that it is independent of application, character, age, gender, ink, paper, pen, etc. Therefore, we can argue that the proposed system tends to generalisation compared to the existing methods [11–13] as the input.

In light of the above discussions, it is observed that most of the methods are developed for specific applications. In addition, none of the methods studies handwritten characters from ‘a’ to ‘z’ for identifying possible personal behaviours. As a result, we can conclude that there is no method which works well irrespective of applications, genders and fields. Therefore, in this work, we propose to study the attributes of characters from ‘a’ to ‘z’ for the identification of personal behaviours. The contributions are as follows: (i) exploring local information of handwritten characters for personal behaviour identification without depending on applications, fields, genders and ages, (ii) deriving hypothesis based on local information and dictionary of graphology for each behaviour of persons, and (iii) an automatic interactive system for generating ground truth and validating the proposed method.

3 Proposed method

It is evident from graphology theory (https://ipip.ori.org/) that personal behaviours are often reflected in their handwriting styles, especially in written loops, stems, height, width and slant of characters with respect to baseline [14]. It is also true that since our intention is to collect datasets online using pen and pad, characters written by different users do not pose any distortion, degradation and noise. This observation motivates us to propose features, such as zone, angle and loop that extract effects and reflections of different personal characteristics of persons. The rationale behind to propose rule-based features is as follows. When images are clean and preserve unique shapes with respect to different behaviours of persons, and graphology provides rules for each behaviour of persons, we believe the proposed features with rules can achieve better results. The proposed rules are used to study shapes and positions of loops, height or angle of stems with respect to baseline, end or branch points, sleeping lines over characters, etc.

Note that the aim of the proposed work is to study personal behaviours (e.g. talkative, broadminded, etc.), positive social (e.g. witty, the ability to make work successful, etc.), negative social behaviours which create a bad social behaviours which create a cool social environment (e.g. irritating, cheating, etc.), and personal behaviour which describes the individual person attitude (e.g. ability, self-confidence, nervousness etc.). To achieve the above-mentioned goal, the proposed method considers hand-written characters and by different users as described in Sec. 1. We propose to extract structural features such as zone-based, loop-based, angular based and other geometrical information of characters for identifying personal behaviours. According to graphologists and with their experiences, we then derive hypothesis using structural features for identifying three types of person behaviours, namely, positive social behaviour, negative social behaviour and personal behaviour. The flow of the proposed method can be seen in Fig. 1.

3.1 Structural feature extraction

In this work, as mentioned in the previous section, we extract different types of structural features, namely, zone-based, loop-based, angular based, stem-based, oval based, branch-based, bar-based etc., for human behaviour identification. For extracting zone-based features, the proposed method traces contours of characters to detect intersection points, holes, concavity and convex hull, which are image processing concepts. Based on geometrical properties of the above shapes, the proposed method divides each whole character into three zones, namely, upper, middle and lower zones, as shown in Fig. 2a. In Fig. 2a, it is noted from character ‘y’ that the intersection point and the hole are helping the method to find lower, middle and upper zones.
In the same way, for character ‘h’, the concavity created in the bottom and the hole created at the top help us to find zones. For a character like ‘c’, since there are no intersection points and branches, the concavity determines the middle zone. For loop-based features, while tracing a boundary, if the proposed method visits the same starting point again, it can be considered as a loop or a hole. For the purpose of finding holes and loops, we have explored the Euler number concept as shown in Fig. 2b. For angle based features, the proposed method uses curvature concept, which estimates angle for the region that have corners, where the contour is cursive as shown in Fig. 2c. For finding ends, intersections and branches of characters, the proposed method uses neighbourhood information along with angle information as shown in Figs. 2d and f. For extracting oval-shaped features, the proposed method then explores the well-known image processing concept called water reservoir model [14], which defines oval based on water collection in the oval area, as shown in Fig. 2e. Similarly, the proposed method explores the run-length smearing concept, which counts successive pixels of the same information for bar detection while tracing the boundary of a character, as shown in Fig. 2g. Overall, the proposed method uses the above-mentioned image processing concepts for feature extraction in this work. Similarly, the proposed method extracts the observations based on loop shapes by estimating height, width, as shown in Fig. 2h. These observations are useful for identifying negative social behaviours. The proposed method observes oval-based angle information as shown in Fig. 2c, which is useful to identify personal behaviours such as ‘the ability to handle critical situations by signing for himself’. Observations are extracted based on the height and width of stems with respect to baseline, as shown in Fig. 2d, which are useful for the three types of behaviours.

Observations based on oval shapes are extracted, as shown in Fig. 2e, where the proposed method uses a water reservoir model [14] for identifying different types of ovals. These observations are useful for all three types of personal behaviours. Observations on end and branch points are extracted using character skeletons as shown in Fig. 2f, which are also useful for identifying the three types of personal behaviours. Note that the proposed method extracts observation based on the bar (sleeping line over the characters), as shown in Fig. 2g, which are useful for identifying the negative social behaviour of the person.

In the same way, we study the characteristics of the above structures of the different characters from ‘a’ to ‘z’ to identify unique observations for different personal behaviours which are quite common and essential to make the system successful in respective fields such as education, medical management, etc. The unique observations and respective behaviour of the person with handwritten characters are listed in Figs. 3–5, respectively, for personal behaviour, positive social and negative social behaviours. To extract the observations from Figs. 3–5, the proposed method derives hypothesis and implement the same to identify the behaviour automatically as conditions are listed in Tables 1–3 for respective observations listed in Figs. 3–5. The acronym and meaning the variables are listed in Table 4 which defines all the variables listed in Tables 1–3.

4 Experimental results

To evaluate the performance of the proposed method, we developed an interactive system which allows users to directly write characters or upload scanned images of handwritten characters. When a user writes or uploads handwritten characters, the system automatically displays responses that represent personal behaviours as listed in Section 3. Afterwards, the user has two options, either reject prediction as ‘I don’t agree with the results’ or accept prediction as ‘I agree with the results’. For each character written by the user, the proposed system predicts user behaviours. If the user agrees with system decision, we count it as one correct. Otherwise, we consider it as a wrong count. In this case, the performance of the system depends on user decisions to agree or disagree.

We believe that the user responds by giving his/her decision to the proposed system without any bias. However, sometimes, it is hard to ensure that user response is genuine for all situations. To overcome this issue, we have collected a new dataset with clear ground truth for experimentation. Since this data provides ground truth of individual person behaviours as the shown sample images and respective ground truth in Fig. 6, we can verify the response given by the user and the proposed system for each instance of writing. In this work, we consider two datasets, namely, dataset-1 without ground truth and dataset-2 with ground truth. Dataset-1 comprises the collection of online handwritten character images and uploaded images, which are offline handwritten character images. For online image collection, we send the link of the proposed system to different users to write character images. This helps us to collect more samples. However, it is hard to verify the response of the user with the output of the proposed system. At the same time, we lose the ground truth of character images.

To overcome this limitation and for a fair evaluation of the proposed method, we create a second dataset (dataset-2) which includes off-line handwritten character images of 30 writers with ground truth as the shown sample images and their ground truth in Fig. 6. In this case, we collect images when the person is present physically. Due to different mechanisms of image collection, one can guess that dataset-1 contains a large variation in the image collection, while dataset-2 contains fewer variations compared to dataset-1. We believe that evaluating the proposed system on different datasets results in fair evaluations for person behaviour identification. More details about dataset-1 and dataset-2 can be found in Fig. 3. Since the primary goal of the proposed work...
Fig. 2 Different structural features for deriving hypothesis

a Zone-based features
b Loop-based features
c Angular based features
d Stem based features
e Oval based features
f Endpoint and branch point-based features
g Bar based features
| NO. | Letter | Handwritten Character | Observations | Behavior |
|-----|--------|-----------------------|--------------|----------|
| 1   | a      |                       | Loop disconnected at right. | Talkative |
| 2   | b      |                       | Upper loop appear as balloon and lower loop has more width than its height. | Broadminded |
| 3   | b      |                       | The upper loop is narrow and long and there is small gap in lower loop. | Follow Idealism and Talkative |
| 4   | b, p, m| Curve within lower loop. Ring like structure at the left of “m” |  | Egoistic |
| 5   | c      |                       | Semi-circle | Socially comfortable |
| 6   | c      |                       | Extra like tangent to contour | Stubborn |
| 7   | c      |                       | Mirror image of “c” | Introvert |
| 8   | d      |                       | Upper loop can be any sized and width of lower loop must be higher than height | Lazy (Wasting time by watching TV) |
| 9   | d      |                       | Width of lower loop should be less than its height. | Restless |
| 10  | d      |                       | Stem with long loop 45 angle deviation from baseline | Sudden Angry and cool down fast |
| 11  | e      |                       | End part of “e” continues with 45 degree angle | Attention seeker |
| 12  | f      |                       | Starting and end of loop at left side | Good in planning but lack of confidence |
| 13  | h      |                       | Narrow loop with long stem | Intelligent and open minded |
| 14  | h      |                       | Wide loop with long stem | Intelligent and broadminded |
| 15  | h      |                       | Loop of Balloon shape with long stem | Intelligent but have confusion |
| 16  | h      |                       | Spreading legs of the character | He loves freedom and does not like to be controlled |
| 17  | h      |                       | Knot at top | Frustration |
| 18  | i      |                       | Cap like a shape on top | Hiding his/her own weaknesses |
| 19  | i      |                       | Dot at upper-right side of character | Advance thinker |
| 20  | i      |                       | There is no dot | Highly stressful personality |
| 21  | m      |                       | Left hump is higher than right hump | Diplomatic |
| 22  | m      |                       | Right hump is greater than left hump | Self-conscious |
| 23  | p      |                       | Disconnection in loop | Talkative |
| 24  | q      |                       | End point is stretched with angle towards up | Aggressive in profession |
| 25  | q      |                       | End point stretches like a hook | Stressful |
| 26  | q      |                       | No stretching at end point | Loner |
| 27  | r      |                       | Angular | Ability to handle/manage |
| 28  | r      |                       | Letter looks like “U” with loop at left side | Sing for herself or himself |
| 29  | s      |                       | Loop end with stretch towards right | Stressful and Stubborn |

Fig. 3 List of hypotheses derived based structural features of handwritten character images for personal behaviour
To measure the performance of the proposed method, we use accuracy, which is defined as the total number of correct responses given by the user (agree or disagree for the proposed system prediction) divided by the total number of characters. To show the usefulness of the proposed method, we implement the state-of-the-art methods, namely, the method [11] which defines rules as the proposed method for human behaviour prediction, and one more method [12] which explores an artificial neural network for human behaviour prediction, for comparative study with the proposed method. The main reason to use these two methods for the comparative study is that these two are the state-of-the-art methods and set the same objective as the proposed method. In the same way, to assess the contribution of the feature extraction and defined rules, we pass the extracted features to CNN and SVM classifier for person behaviour identification. The number of samples for training and testing for the dataset-1 and dataset-2 is listed in Table 5. The same set up is used for all the experiments in this work. For SVM classifier, we follow the instructions in [15] and for CNN, we use the pre-defined architecture as in [16] for experimentation.

For implementing CNN, we use the architecture proposed in the method [16] for image recognition, which is called VGG architecture. The details of the architecture are as follows. VGG architecture involves layers in the first, second, third, fourth and fifth convolution blocks containing 64, 128, 256, 512 filters, respectively. All the layers use filter size of 3 × 3 with ReLU activation. After the fifth convolution block, we flatten the output and add three dense layers with 1024, 1024 and 59 nodes, respectively. The first and second dense layers have ReLU activation, whereas the third dense layer has Softmax activation. We use 0.5 dropout between these two dense layers to reduce chances of overfitting. The model contains 16 M trainable parameter. We train the network using SGD optimiser with learning rate 0.0001, momentum 0.5 and batch size 50. More details can be found in [16].

4.1 Evaluating the proposed person behaviour identification method

The accuracies of the proposed features + rules, CNN, SVM and existing methods are reported in Table 6, where it is noted that the proposed features + rules achieve the best accuracy compared to all the other methods. It is observed from Table 6 that the proposed and existing methods score better results for dataset-2 compared to dataset-1. This is due to more samples with large variations in image collection compared to dataset-2. It is also true that writing on paper using pens gives more natural writing of individuals than writing using online devices. When we compare the proposed features + rules with the proposed features + CNN and the proposed features + SVM, the proposed features + rules are better than the other two methods for both dataset-1 and dataset-2 as reported in Table 6. The reason is that the proposed method derives rules for each person behaviour identification, according to graphologist, and the rule are unique in nature and do not overlap with other rules. Moreover, the proposed shape-based features are invariant to different variations caused by pen, paper, ink, device, online writing, off-line writing, etc. It is evident from the results shown in Fig. 8a, where it is seen that as long as the structure of a character is preserved, the proposed rules work well.

Table 6 shows that the proposed features with SVM are better than the proposed features with CNN. This is valid because of freestyle writing as one can expect large variations in writing. Besides, since users use a special pen and pad for writing characters, which is not the usual practice of writing, we can expect more variations than writing on paper. When we have large variations, CNN requires more samples to achieve better results as it helps to

| NO. | Letter | Handwritten Character | Observations | Behaviour |
|-----|--------|-----------------------|--------------|-----------|
| 1   | k      | K                     | Regular Character | Person is witty |
| 2   | k      | K                     | Closed loop at left side of the stem | Witty and Romantic |
| 3   | th     | th                    | Only one bar for both the character | Has extra mental ability to make work successful |
| 4   | y      | y                     | Huge curb at lower zone | Love with love |

Fig. 3 (Continued)

Fig. 4 List of hypothesis derived based structural features of handwritten character images for the positive social behaviour of the person

| Handwritten Character | Observations | Behaviour |
|-----------------------|--------------|-----------|
| Step and bar are almost straight | Rigid |
| Starting or ending strokes look line. | Glamorous |
| Regular V, Regular Z | Analytical Person |
| Left and right strokes are perpendicular to base line. | Independent |
| Loop end stretches towards right like straight line | Rigid, frustration and fear |
| Two Wavy bars on top and bottom for ‘T’, Wavy line for ‘V’, Wavy stroke at the end for ‘Z’. | Fun loving person |

is to predict person behaviours irrespective of age, gender, sex, qualification, paper, pen, ink, application, etc. dataset-1 and dataset-2 do not provide the above details.

Fig. 7 shows the sample screenshots of the proposed interactive system, where we can see ‘upload’ option to get handwritten characters by scanning, ‘clear’ option to remove the previous results displayed on screen, ‘analyse’ option to send request to the system to test characters, ‘I don’t agree with the result’ option is to reject the response given by the system for the written characters, and ‘I agree with the result’ option is to accept the responses given by the system for the characters. Sample response for character ‘a’ written by the user is displayed as ‘the respective person is talkative’. This is one sample of a particular person.

This is an open access article published by the IET, Chinese Association for Artificial Intelligence and Chongqing University of Technology under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
determine the proper weight for classification. However, collecting a large number of samples for this work is hard because people do not like to share their personal behaviours. On the other hand, SVM does not require more samples for classification if a feature extracts the

\[ \text{Table 1 Conditions for the hypotheses listed in Fig. 3} \]

| No | Equations for the hypotheses | No | Equations for the hypotheses |
|----|-------------------------------|----|-------------------------------|
| 1  | \( \text{NWR}_R = 1 \) | 19 | \( H = 0 \) & \( \Delta_i (L) < \text{MAX}(L_i(S)) \) |
| 2  | \( L_h < 3' L_w \& O_h < O_w \) | 20 | \( I_b = 0 \& H = 0 \) |
| 3  | \( \text{NWR}_R = 0 \& L_h < 2' L_w \) | 21 | \( L_{HM_s} > R_{HM_s} \) |
| 4  | \( \text{NWR}_b = 3 \) for m \( \text{NWR}_b = 2 \) for b \( \text{NWR}_b = 3 \) for p | 22 | \( L_{HM_s} < R_{HM_s} \) |
| 5  | \( \text{NWR}_b = 3 \) | 23 | \( \text{NWR}_R = 1 \) |
| 6  | \( \text{NWR}_R = 1 \) | 24 | \( \Delta_i (O) < \Delta_i (E_{point}) \) |
| 7  | \( \text{NWR}_R = 1 \) | 25 | \( H = 1 \) |
| 8  | \( O_h < O_w \) | 26 | \( H = 0 \& D = 1 \& L = 0 \) |
| 9  | \( O_h > O_w \) | 27 | \( A_{point} = 0 \) |
| 10 | \( S_w = 45' \) | 28 | \( A_{point} = 1 \) |
| 11 | \( \alpha = \text{the angle between} (E_{point}, B_{point}) \& \beta \approx 45' \) \( \alpha < \beta \) \( \alpha = 10' \) | 29 | \( \Delta_i (E_{point}) = \text{MAX}(\Delta_i(S)) \) |
| 12 | \( \Delta_i (E_{point}) = \text{MAX}(\Delta_i(S)) \) | 30 | \( S_w \) is constant in every row |
| 13 | \( L_h > 3' L_w \) | 31 | \( \text{NWR}_R = 2 \) |
| 14 | \( S_h = 0.6' C_h \& L_h \) | 32 | \( L = 0 \& \text{NWR}_R = 1 \) for v \( L = 0 \& \text{NWR}_R = 2 \) for w |
| 15 | \( S_h = 0.6' C_h \& L_h \) | 33 | \( L = 0 \& O = 0 \& \text{NWR}_R = 3 \) |
| 16 | \( O_w > O_h \) | 34 | \( L = 1 \& L_h \approx 0.4' C_h \) |
| 17 | \( L_h < 0.25' C_h \) | 35 | \( \text{NWR}_R > 3 \) |
| 18 | \( H = 1 \) | — | — |

\[ \text{Table 2 Conditions for the hypotheses listed in Fig. 4} \]

| No | Equations for the hypotheses | 3 | \( L = 1 \) |
|----|-------------------------------|----|-------------------------------|
| 1  | \( \text{NWR}_b = 3 \) | 3 | \( L = 1 \) |
| 2  | \( O = 1 \& \text{NWR}_b = 4 \) | 4 | \( \text{Curb}_w < \text{curb}_w \) |

Fig. 5 List of hypothesis derived based on contours of handwritten character images for the negative social behaviour of the person

Table 2 Conditions for the hypotheses listed in Fig. 4

| No | Equations for the hypotheses | 3 | \( L = 1 \) |
|----|-------------------------------|----|-------------------------------|
| 1  | \( \text{NWR}_b = 3 \) | 3 | \( L = 1 \) |
| 2  | \( O = 1 \& \text{NWR}_b = 4 \) | 4 | \( \text{Curb}_w < \text{curb}_w \) |

This is an open access article published by the IET, Chinese Association for Artificial Intelligence and Chongqing University of Technology under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
Table 3 Conditions for the hypotheses listed in Fig. 5

| Condition | Expression |
|-----------|------------|
| 1 | \( O = 1 \& L_B > 2''d_P \) |
| 2 | \( (E_P \& B_P) \) makes a straight line |
| 3 | \( \Delta \left( E_{point} \right) < \Delta \left( B \right) \) |
| 4 | \( E_{point} = x \left( L_B \right) = E_{point} \left( d_P \right) < 0.2'' C_m \) |
| 5 | \( \Delta \left( B \right) < 0.2'' C_m \) |
| 6 | \( \Delta \left( E_{point} \right) > \Delta \left( M_e \right) \) |
| 7 | \( \Delta \left( B \right) < \Delta \left( B \right) \) |
| 8 | \( \Delta \left( B \right) > \Delta \left( B \right) \) |
| 9 | \( \Delta \left( B \right) > \Delta \left( B \right) \) |
| 10 | \( \Delta \left( B \right) > \Delta \left( B \right) \) |

Table 4 Acronyms used for hypotheses derivation listed in Tables 1-3

| Acronym | Description |
|---------|-------------|
| C | character height |
| L | loop height |
| S | stem height |
| O | oval height |
| WR | number of water reservoir from bottom |
| WRP | number of water reservoir from right |
| HM | left hump height of m |
| P | end point |
| CRB | curb height |
| Z | lower zone |
| B | bar left point |
| BR | bar height |
| D | coordinate of the point |
| EP | right most point of e |
| BP | lower most point of e |
| CW | character width |
| LW | loop width |
| SW | stem angle with baseline |
| OW | oval Width |
| NWR | number of water reservoir from top |
| RHM | right hump height of m |
| PB | branch point |
| CRB | curb width |
| BM | middle zone |
| BR | bar right point |
| BW | bar height |
| H | hat |

Fig. 6 Sample images of dataset-2 and their ground truth

Fig. 7 Screenshot of the proposed interactive system to validate the responses against character written by users

Table 5 Statistical details of the dataset for evaluation

| Dataset Type | Total Number of Writers | Total Number of Training Samples | Total Number of Testing Samples | Total Number of Samples |
|--------------|-------------------------|---------------------------------|---------------------------------|-------------------------|
| Dataset-1    | 600                     | 3700                            | 900                             | 4600                    |
| Dataset-2    | 30                      | 500                             | 200                             | 700                     |
test the proposed method on capital letters, we pass both lower and upper case letters to the system, as shown in Fig. 8a, where it can be seen that the proposed method predicts the same behaviour for both lower and upper case letters, ‘W’. Since the proposed system extracts features based on character shapes to predict person behaviours, as long as the structure of a character is preserved, the proposed system works well as shown in Fig. 8a. If the shapes and structures of upper and lower case letters are different, the proposed method predicts different behaviours as shown in Fig. 8b, where one can see different behaviours for upper and lower cases of the same character. This is true because when shape and structure change, the rule for person behaviour also changes according to graphology.

Similarly, the proposed method is also tested on noisy images to assess performance. Sample images with added Gaussian noises manually at different levels are shown in Fig. 9, where one can see as Gaussian noise level increases, noise density (the number of Gaussian noise pixels) increases. For images with Gaussian noises, we calculate accuracy using the proposed feature with rules, the proposed features with SVM, and the proposed features with CNN, as shown in Fig. 10. It is noted from Fig. 10 that the proposed features with rules, SVM and CNN do not work well for noise images as we can see the performance decrease as noise

### Table 6  Accuracy of the proposed and existing methods for both the datasets (in %)

| Datasets  | Proposed + Rule | Proposed + CNN | Proposed + SVM | [11] Proposed + Rule | [12] Specific characters ‘γ’ |
|-----------|-----------------|----------------|----------------|----------------------|-----------------------------|
| Dataset-1 | 69.95           | 61.60          | 67.25          | 51.0                 | 99.95                       |
| Dataset-2 | 86.70           | 69.55          | 78.75          | 53.0                 | 92.70                       |

**Fig. 8** Effectiveness of the proposed system for upper case and lower case letters

a Proposed system predicts the same behaviour for both upper case and lower case letters

b Proposed system predicts different behaviour for upper case and lower case letters

**Fig. 9** Sample images affected by Gaussian noise at different levels

0.0 0.2 0.4 0.6 0.8 0.9

**Fig. 10** Performance decrease as noise increases

This is an open access article published by the IET, Chinese Association for Artificial Intelligence and Chongqing University of Technology under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)
written character features do not match with the corresponding character.

5 Conclusions and future work

We have proposed an automatic system for identifying human behaviours based on handwriting at character level by keeping real-time applications of graphology. The proposed method extracts structural features such as slant, hole, aspect ratio, height, width and shape of written characters. According to graphology, the proposed method derives hypothesis from identifying personal behaviours. In this work, we consider possible personal behaviours for identification. To validate the proposed method, we have developed an interactive automatic system to test the hypothesis, which accepts written characters from different persons to predict specific behaviours. The system also allows users to choose to accept or reject each predicted behaviour. Experimental results and a comparative study on two datasets show that the proposed method outperforms the existing methods in terms of accuracy. Furthermore, we can also conclude that the proposed features + rules achieve the best accuracy compared to the proposed features + CNN and the proposed features + SVM.

When a person writes touching characters, the performance of the proposed system degrades because it accepts individual characters as the input for person behaviour identification in this work. Therefore, we have planned to combine character segmentation from the handwritten text line and person behaviour identification in the future. Sometimes, when a person writes characters with more variations in shape, the extracted features may overlap with features of other characters. This leads to poor results for the proposed system. Therefore, the proposed system requires high-level features, such as context between successive characters in words to improve the results. It is true that the rules defined based on graphologist are limited to the number of person behaviours. In order to improve the performance of the proposed method for overcoming the above-mentioned limitations, we plan to explore different CNN architectures using online time information of the writings in the near future.

6 Acknowledgments

The work described in this paper was supported by the Science Foundation for Distinguished Young Scholars of Jiangsu under grant no. BK201606021, the Natural Science Foundation of China under grant nos. 61672273 and 61272218. The authors acknowledge the help of Mr Aishik Konwer and Mr Abir Bhownick of Department of ECE, Institute of Engineering & Management, Kolkata in this work.

7 References

[1] Tett, R.P., Palmer, C.A.: ‘The validity of handwriting elements in relation to self-report personality trait measures’, vol. 22 (Person Individual Difference, UK, 1997), pp. 11–18
[2] Brewer, J.F.: ‘Graphology’, vol. 5, ‘Complementary therapies in nursing & midwifery’, (Royal College of Nursing Publishing Company (RCN), UK, 1999), pp. 6–14
[3] Hafemann, L.G., Sabourin, R., Oliveira, L. S.: ‘Analyzing features learned for offline signature verification using deep CNNs’. Proc. ICPR, Mexico, 2016, pp. 2989–2994
[4] Hurtado, O.M., Guest, R., Stevenage, S.V., et al.: ‘The relationship between handwritten signature production and personality traits’. Proc. IJCB, USA, 2014, pp. 1–8
[5] Gattupalli, V., Chandakkar, P.S., Li, B.: ‘A computational approach to relative aesthetics’. Proc. ICPR, Mexico, 2016, pp. 2446–2451
[6] Majumdar, A., Krishnan, P., Jawahar, C. V.: ‘Visual aesthetic analysis for handwritten document images’. Proc. ICFHR, China, 2016, pp. 423–428
[7] Kodat, S.V., Rommane, D.S.: ‘Automatic personality assessment: a systematic review’. Proc. ICIP, Canada, 2015, pp. 326–331
[8] Fallah, B., Khotanlou, H.: ‘Identify human personality parameter based on handwriting using neural network’. Proc. IRNOPEN, Canada, 2016, pp. 120–126

This is an open access article published by the IET, Chinese Association for Artificial Intelligence and Chongqing University of Technology under the Creative Commons Attribution License

(http://creativecommons.org/licenses/by/3.0/)
[9] Asra, S., Shubhangi, D.C.: ‘Human behavior recognition based on handwritten cursive by SVM classifier’. Proc. ICEECCOT, Iran, 2017, pp. 260–268
[10] Topaloglu, M., Ekmecki, S.: ‘Gender detection and identifying one’s handwriting with handwriting analysis’. Expert Syst. Appl., 2017, 79, pp. 236–243
[11] Champa, H.N., Komar, K.R.A.: ‘Artificial neural network for human behavior prediction through handwriting analysis’, Int. J. Comput. Appl., 2010, 2, pp. 36–41
[12] Champa, H.N., Kumar, K.R.A.: ‘Automated human behavior prediction through handwriting analysis’. Proc. ICIIC, India, 2010, pp. 160–165
[13] Coll, R., Frunes, A., Llados, J.: ‘Graphological analysis of handwritten text documents for human resource recruitment’. In Proc. ICDAR, Australia, 2009, pp. 1081–1085
[14] Pal, U., Belaid, A., Choisy, C.: ‘Touching numeral segmentation using water reservoir concept’, Pattern Recognit. Lett., 2003, 24, pp. 261–272
[15] Boser, B.E., Guyon, I.M., Vapnik, V.N.: ‘A training algorithm for optimal margin classifiers’. Proc. COLT, USA, 1992, pp. 144–152
[16] Simonyan, K., Zisserman, A.: ‘Very deep convolutional networks for large-scale image recognition’. Proc. ICLR, USA, 2015, pp. 1–14