

Artificial Neural Network for Human Behavior Prediction through Handwriting Analysis

Champa H N

Assistant Professor,

Department of Computer Science & Engg.,
University Visveswaraya College of Engineering,
Bangalore-560 001, Karnataka, INDIA

Dr. K R AnandaKumar

Professor & Head,

Department of Computer Science & Engg.,
SJB Institute of Technology, Uttarahalli road, Kengeri,
Bangalore-560 060, Karnataka, INDIA

ABSTRACT

Handwriting Analysis or Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and many others.

Professional handwriting examiners called graphologist often identify the writer with a piece of handwriting. Accuracy of handwriting analysis depends on how skilled the analyst is. Although human intervention in handwriting analysis has been effective, it is costly and prone to fatigue. Hence the proposed methodology focuses on developing a tool for behavioral analysis which can predict the personality traits automatically with the aid of a computer without the human intervention.

In this paper a method has been proposed to predict the personality of a person from the baseline, the pen pressure and the letter 't' as found in an individual's handwriting. These parameters are the inputs to the Artificial Neural Network which outputs the personality trait of the writer. The performance is measured by examining multiple samples.

Keywords

Handwriting Analysis, Personality Traits, Polygonalization, behavioral analysis, Artificial Neural Network

1. INTRODUCTION

Handwriting Analysis or Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Handwriting reveals the true personality including emotional outlay, fears, honesty, defenses and over many other individual personality traits. Handwriting Analysis is not document examination, which involves the examination of a sample of handwriting to determine the author. Handwriting is often referred to as brain writing. Each personality trait is represented by a neurological brain pattern. Each neurological brain pattern produces a unique neuromuscular movement that is the same for every person who has that particular personality trait. When writing, these tiny movements occur unconsciously. Each written movement or stroke reveals a specific personality trait. Graphology is the science of identifying these strokes as

they appear in handwriting and describe the corresponding personality trait.

In this paper a method has been proposed to predict the personality of a person from the features extracted from his handwriting using Artificial Neural Networks. The personality traits revealed by the baseline, the pen pressure and the letter 't' as found in an individual's handwriting are explored in this paper. Three parameters, the baseline, the pen pressure and the height of the t-bar on the stem of the letter 't' are the inputs to the ANN which outputs the personality trait of the writer. The evaluation of the baseline is using the polygonalization method and the evaluation of the pen pressure utilizes the grey-level threshold value. The height of the t-bar on the stem of the letter 't' is calculated using template matching. The baseline, the pen pressure and the letter 't' in one's handwriting reveal a lot of accurate information about the writer. MATLAB is the tool used for the purpose. The performance is measured by examining multiple samples.

2. RELATED WORK

Handwriting Analysis or Graphology is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Among the many aspects of handwriting that can serve as scheme to predict personality traits are baseline, size of letters, writing pressure, connecting strokes, spacing between letters, words and lines, starting strokes, end-strokes, word-slant, speed of handwriting, width of margins, and others [1][2] [3]. Writer individuality rests on the hypothesis that each individual has consistent handwriting, which is distinct from the handwriting of another individual. However, this hypothesis has not been subjected to rigorous scrutiny with the accompanying experimentation, testing, and peer review [4] [5] [6] [7] [8] [9].

3. PROPOSED METHODOLOGY

Professional handwriting examiners called graphologist often identify the writer with a piece of handwriting. Accuracy of handwriting analysis depends on how skilled the analyst is. Although human intervention in handwriting analysis has been effective, it is costly and prone to fatigue. Hence the proposed methodology focuses on developing a tool for behavioral analysis which can

predict the personality traits automatically with the aid of a computer without the human intervention.

The most common feature of handwriting that can serve as scheme to predict personality traits are baseline, size of letters, writing pressure, connecting strokes, spacing between letters, words and lines, starting strokes, end-strokes, word-slant, speed of handwriting, width of margins, and others. In this paper the baseline, the writing pressure and the height of the t-bar on the stem of the letter 't' are considered for predicting the personality of the writer[10].

The baseline and the pen pressure in one's handwriting reveals a lot of accurate information about the writer. Baseline in one's handwriting is the line along which the writing flows. The three most common baselines found in any handwriting are **ascending, descending and level** as shown in Fig.1. Appropriately each one relates to a specific personality trait of the writer.

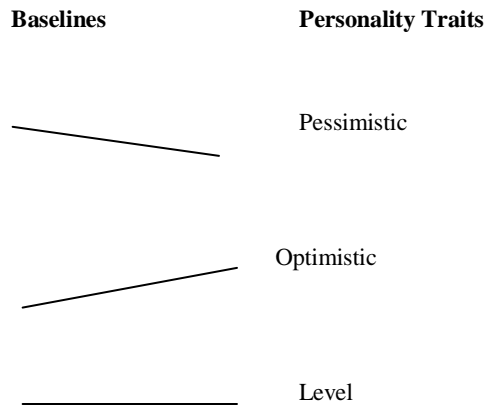


Figure 1: The baselines representing different personality traits of the writer

The other most important feature in a handwriting is the **writing pressure**. The amount of pressure exerted on the paper while writing is the **depth of feeling**, also called **emotional intensity**. Based on the pen pressure the writer can be classified as **light writer, medium writer** and **heavy writer**. Table-1. shows the corresponding personality traits of the different pen pressures.

Table 1. Personality traits represented by different pen pressures.

Writer category	Personality trait
Light writer	This writer can endure traumatic experiences without being seriously affected. Emotional experiences do not make a lasting impression on him.
Medium writer	This writer feels emotional experiences for a moderate amount of time. He has an average level of emotional intensity.
Heavy/deep writer	This person has very deep and enduring feelings. This writer may forgive, but he will never forget. He feels situations intensely.

The main technique for finding the baseline slant is the polygonalization of the single line in the handwritten text[11]. The pen pressure can be calculated as a count of the number of foreground pixels in the thresholded image. The number of black pixels is indicative of the pen-pressure, thickness of strokes, and the size of writing[13].

The lower case letter 't' is one of the letters in one's handwriting that reveals a lot of accurate information about the writer. People write letter 't' in many different ways. There are various ways to make the stem, the cross on the t-bar, and even the entry and exit to this letter 't', each one of which relates to a specific personality trait of a person. Fig.2 shows some of the ways in which letter 't' is written.

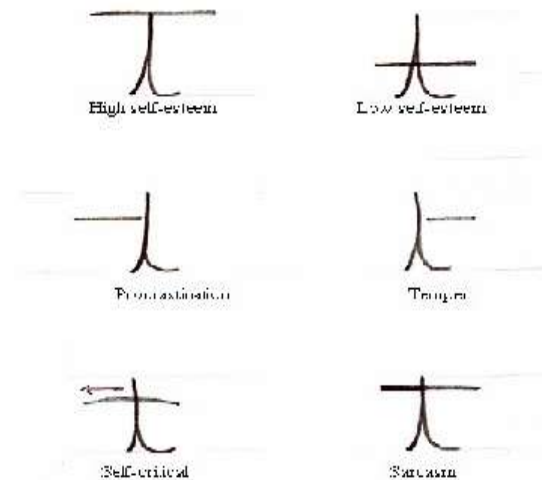


Figure 2: The scanned patterns of letter 't' representing different traits of the writer

One of the personality trait revealed by letter 't' is the **self-esteem** of the writer. When the t-bar is crossed very high (not above the stem), it indicates high self-esteem. Good self-confidence, ambition and goals are

indicated here. The t-bar crossed very low on the stem indicates low self-esteem. This person fears failure and resists change. The main technique for automating the character recognition is the template matching.

The methodology of identification of certain traits of the writer through examining the letter 't' involves template matching with certain predefined templates [10]. The predefined templates are created each of which represent a particular personality. It is incorporated as a measure of how well an arbitrary pattern of grey values, a *template* $g(x,y)$, matches a given image $f(x,y)$ [14]. The minima of these measures are the best match. Five templates are created for the letter 't' with a 6x5 grid. The five templates are for the five different positions of the t-bar on the stem. Each of the grids represents a certain position of the t-bar and accordingly certain character of the writer. The handwriting sample is scanned and the image is taken as the input. Only the letter 't's in the image are selected. These are resized to fit into the 6x5 grid. Some

of the personality traits that are revealed by the t-bar of the letter 't', considered in this paper are listed in table 2.

The decision making on the matched pattern is achieved by hamming distance [15]. The Hamming distance between two strings of equal length is the number of positions for which the corresponding symbols are different [15]. Put another way, it measures the minimum number of *substitutions* required to change one into the other, or the number of *errors* that transformed one string into the other. This application considers only binary patterns. For binary strings a and b the Hamming distance is equal to the number of ones in a XOR b .

Table 2: Personality traits revealed by letter 't'

Sl.No	Position of the t-bar	Personality trait
1.	Crossed very high (not above the stem, but in a very high region)	High self-esteem: This reveals confidence, ambition, and the ability to plan ahead, high goals, high personal expectations and an overall good self-image. This is the key to personal success and happiness.
2.	Crossed above the middle zone but not at the loop	Moderate self-esteem: This person is practical. This is a positive trait and common among successful people.
3.	Crossed very low on the stem	Low self-esteem: This person fears failure and resists change. He often remains in bad situations and relationships for too long and finds imperfections with himself. He is rarely successful enough in his own eyes despite his accomplishments.
4.	Crossed above the stem	Dreamer : This person's goals and dreams lose touch with reality. These people often talk about what they are going to do instead of doing it.

4. IMPLEMENTATION & RESULT

The values calculated for the height of the t-bar on the stem of the letter 't' (5 different positions), the pen pressure(2 values for dark and light writing), the baseline slant(3 values for three different slants) are taken as the inputs to

the ANN. The 30 different outputs of the ANN are the 30 different personality traits of the writer depending on the way he/she crosses his t, the pressure with which he/she writes and the baseline of his/her writing. The huge amount of data amounting from 30 different combinations of traits is to be analyzed to predict the personality traits. The involvement of enormous feature data prompts for use of artificial neural network for automated prediction.

Back propagation is the generalization of the Widrow-Hoff learning rule to multiple-layer networks and nonlinear differentiable transfer functions [16]. Input vectors and the corresponding target vectors are used to train a network until it can approximate a function, associate input vectors with specific output vectors, or classify input vectors in an appropriate way as already defined.

Standard back propagation is a gradient descent algorithm, similar to the Widrow-Hoff learning rule, in which the network weights are moved along the negative of the gradient of the performance function. The term *back propagation* refers to the manner in which the gradient is computed for nonlinear multilayer networks. Properly trained back propagation networks tend to give reasonable answers when presented with inputs that they have never trained with. Typically, a new input leads to an output similar to the correct output for input vectors used in training that are similar to the new input being presented. This generalization property makes it possible to train a network on a representative set of input/target pairs and get good results without training the network on all possible input/output pairs.

The back propagation neural network is a multilayer network that comprises of input, hidden and output layers. The number of nodes in the input

layer is equal to the number of features used for class representation. The number of nodes in the output layer is equal to the number of classes needed. The number of nodes in the hidden layer mainly decides the refinement to the inputs such that the output goal is reached in shortest duration of training period.

The mathematical concepts governing the back propagation neural network are governed by the basic algorithm comprising of the basic steps as,

Initialize all weights to small random numbers.

Until satisfied, Do

For each training example, Do

1. Input the training example to the network and compute the network outputs
2. For each output unit k
 $\delta_k \leftarrow O_k(1-O_k)(t_k-O_k)$
3. For each hidden unit h
 $\delta_h \leftarrow O_h(1-O_h) \sum W_{h,k} \delta_k$
4. Update each network weight $w_{i,j}$

$$w_{i,j} \leftarrow w_{i,j} + \Delta w_{i,j}$$

where $\Delta w_{i,j} = \eta \delta_j x_{i,j}$

For output units

Calculate $\frac{\partial E}{\partial w_{ji}}$ for each input weight W_{ji} for each output unit j. Note first that since Z_j is a function of W_{ji}

$$\begin{aligned} \frac{\partial E}{\partial w_{ji}} &= \frac{\partial E}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ji}} \\ &= \frac{\partial E}{\partial z_j} x_{ji} \end{aligned}$$

$$E = \frac{1}{2} \sum_{k \in \text{Outputs}} (t_k - \sigma(z_k))^2$$

The outputs of all units $K \neq J$ are independent of W_{ji} , we can drop the summation and consider the contribution to E by j.

$$\begin{aligned} \delta_j = \frac{\partial E}{\partial z_j} &= \frac{\partial}{\partial z_j} \frac{1}{2} (t_j - o_j)^2 \\ &= -(t_j - o_j) \frac{\partial o_j}{\partial z_j} \\ &= -(t_j - o_j) \frac{\partial}{\partial z_j} \sigma(z_j) \\ &= -(t_j - o_j)(1 - \sigma(z_j))\sigma(z_j) \\ &= -(t_j - o_j)(1 - o_j)o_j \end{aligned}$$

For output units

$$\Delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \delta_j x_{ji}$$

For Hidden units

We want to calculate $\frac{\partial E}{\partial w_{ji}}$ for each input weight W_{ji} for each hidden unit j. Note that W_{ji} influences just Z_j which

influences O_j which influences $\forall k \in \text{Downstream}(j)$ each of which influence E. So we can write,

$$\begin{aligned} \frac{\partial E}{\partial w_{ji}} &= \sum_{k \in \text{Downstream}(j)} \frac{\partial E}{\partial z_k} \cdot \frac{\partial z_k}{\partial o_j} \cdot \frac{\partial o_j}{\partial z_j} \cdot \frac{\partial z_j}{\partial w_{ji}} \\ &= \sum_{k \in \text{Downstream}(j)} \frac{\partial E}{\partial z_k} \cdot \frac{\partial z_k}{\partial o_j} \cdot \frac{\partial o_j}{\partial z_j} \cdot x_{ji} \end{aligned}$$

Again note that all the terms except x_{ji} in the above product are the same regardless of which input weight of unit j we are trying to update

Also:

$$\frac{\partial E}{\partial z_k} = \delta_k \cdot \frac{\partial z_k}{\partial o_j} = w_{kj} \text{ and } \frac{\partial o_j}{\partial z_j} = o_j(1 - o_j).$$

So:

$$\begin{aligned} \delta_j &= \sum_{k \in \text{Downstream}(j)} \frac{\partial E}{\partial z_k} \cdot \frac{\partial z_k}{\partial o_j} \cdot \frac{\partial o_j}{\partial z_j} \\ &= \sum_{k \in \text{Downstream}(j)} \delta_k w_{kj} o_j (1 - o_j) \end{aligned}$$

The number of nodes in the hidden layer mainly decides the refinement to the inputs such that the output goal is reached in shortest duration of training period. We therefore verify the time lapsed in training session for multiple number of hidden layer nodes for the outputs to reach a value that is nearest to the predefined target output by a factor of 0.1. The results indicate that the performance goal is reached when the number of hidden layer nodes is 8 and the number of epochs is 4500 in the shortest duration of time. Decrease in the value of both hidden layer nodes and epochs results in under training and more number in over training. Table 3 shows the goal values for different number of hidden layer nodes. The training performance graph is shown in Fig 3.

Table 3: The Performance goal values for different number of hidden layer nodes.

Hidden layer nodes						
	02	04	06	08	10	12
Epochs	Performance					
500	0.53	0.42	0.34	0.35	0.36	0.27
1000	0.36	0.35	0.29	0.31	0.48	0.22
1500	0.32	0.34	0.28	0.29	0.41	0.51
2000	0.22	0.22	0.25	0.25	0.53	0.31
2500	0.21	0.2	0.24	0.16	0.23	0.31
3000	0.16	0.19	0.22	0.12	0.2	0.25
3500	0.12	0.15	0.13	0.15	0.18	0.19
4000	0.16	0.14	0.15	0.13	0.17	0.14
4500	0.15	0.14	0.17	0.099	0.11	0.16
5000	0.13	0.11	0.099	0.099	0.099	0.11
5500	0.11	0.099	0.099	0.099	0.099	0.099
6000	0.099	0.099	0.099	0.099	0.099	0.099



Figure 3: The training performance graph

5. CONCLUSION

A method has been proposed to predict the personality of a person from the features extracted from his handwriting using Artificial Neural Networks. The personality traits revealed by the baseline, the pen pressure and the letter 't' as found in an individual's handwriting are explored in this paper. Three parameters, the baseline, the pen pressure and the height of the t-bar on the stem of the letter 't' are the inputs to the ANN which outputs the personality trait of the writer. The evaluation of the baseline is using the polygonalization method and the evaluation of the pen pressure utilizes the grey-level threshold value. The height of the t-bar on the stem of the letter 't' is calculated using template matching.

This paper proposes an off-line, writer independent handwriting analysis system. Even though offline handwriting recognition lacks the interactive nature of on-line handwriting recognition, this is taken care in this work by allowing the user to select the required character from the input image for feature extraction. The future work can be to include more features of the handwriting like the size of the letters, the margins and others as inputs for personality trait determination.

6. REFERENCES

- [1] Champa H N, K R AnandaKumar, "A Scientific Approach to Behavior Analysis through Handwriting Analysis", National Conference on Research Trends in Information Technology, S R K R Engineering College, Bhimavaram, Andhra Pradesh, 2008
- [2] N Mogharreban, S Rahimi, M Sabharwal, "A Combined Crisp and Fuzzy Approach for Handwriting Analysis", IEEE Annual Meeting of the Fuzzy Information, 2004, vol1, pp 351-356
- [3] Sung-Hyuk and Charles C Tappert, "Automatic Detection of Handwriting Forgery", Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, 2002, vol1, pp.351-356

- [4] K Toraichi, T Horiuchi, R Haruki, “**Observation Method for Mathematical Graphology**”, Proceedings of the Third International Conference on Document Analysis and Recognition, 1995, vol2, pp.656-659.
- [5] Sung-Hyuk Cha, Sargur N Srihari, “**Apriori Algorithm for Sub-category Classification Analysis of handwriting**”, Proceedings of the Sixth International Conference on Document Analysis and Recognition, 2001, pp. 1022-1025.
- [6] H.E.S.Said, T.N.Tan and K.D.Baker, “**Writer Identification Based on Handwriting**”, IEE Third European workshop on Handwriting Analysis and Recognition, vol.33, no.1, 2000, pp 133-148.
- [7] Philip Jonathan Sutanto, Graham Leedham and Vladimir Pervouchine, “**Study of the Consistency of some Discriminatory Features used by Document Examiners in the Analysis of Handwritten letter ‘a’**”, Proceedings of the seventh International Conference on Document Analysis and Recognition, 2003.
- [8] Saragur N Srihari and Zhixin Shi, “**forensic Handwritten Document Retrieval System**”, Proceedings of the First International Workshop on Document Image Analysis for Libraries (DIAL’04), 2004.
- [9] Ameer BENSEFIA, Ali NOSARY, Thierry PAQUET, Laurent HEUTTE, “**Writer Identification by Writer’s Invariants**”, Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition”, 2002, pp 274-279.
- [10] Champa H N, K R AnandaKumar, “**Handwriting Analysis for Writer’s Personality Prediction**”, International Conference on Biometric Technologies and Applications- the Indian perspective, Under the aegis of BiometricsIndia Expo 2008, New Delhi, India, pp 182-191
- [11] Champa H N, K R AnandaKumar, “**Rule Based Approach for Personality Prediction Through Handwriting Analysis**”, 2nd International Conference on Biomedical Informatics and Signal processing, organized by Sai’s BioSciences Research Institute Pvt. Ltd., 2009
- [12] Champa H N, K R AnandaKumar, “**Writer’s Personality Prediction through letter ‘y’ using Generalized Hough Transform(GHT)**”, 3rd International Conference on Information Processing, organized by the Society of Information Processing, Bangalore, Defence Institute of Advanced Technology and University Visveswaraya College of Engineering, Bangalore, 2009
- [13] Srihari S.N., Sung-Hyuk Cha and Sangjik Lee, “**Establishing handwriting Individuality using pattern recognition techniques**”, Proceedings of the Sixth International Conference on Document Analysis and Recognition, 2001, pp. 1195 – 1204.
- [14] Anlong Ming, et, al, “**A Grid-based Face Recognition Approach Using General Template Matching**”, Proceedings of the First International Conference on Semantics, Knowledge, and Grid ,IEEE 2006.
- [15] Hui Tian and Hong Shen, “**Hamming Distance and Hop Count Based Classification for Multicast Network Topology Inference**”, Proceedings of the 19th International Conference on Advanced Information Networking and Applications (AINA’05).
- [16] Mohammad Ali Lotfollahi, et, al, "A New Method in Determining Random Wave-Induced Inline Force", World Applied Sciences Journal, Vol 3 (4), pp 674-683, 2008.