Relationship between personality and handwriting of Chinese characters using artificial neural network

Tingting Wang, Zhanghui Chen*, Wenlong Li, Xiaohui Huang, Pengfei Chen, Siyao Zhu
Department of Electronic Science and Technology, Huazhong University of Science and Technology
1037 Luoyu Road, Wuhan, China
*zhanghuichen88@gmail.com

***Abstract*—This paper adopts artificial neural network to identify the relationship between personality and handwriting of Chinese characters. The handwriting feature studied here is character spacing, which is extracted by samples scanning and image processing. And the personality factors are obtained by Cattell's 16PF. Results of 16-8-1 neural network show that the association is significant between 16 personality factors and character spacing. And from the training and non-linear fitting of 16-1-1 network, which is trained alternately by genetic algorithm and Levenberg-Marquardt algorithm, it is found that character spacing has strong correlations with reasoning and sensitivity.**

Keywords: graphology; handwriting of Chinese characters; Cattell's 16PF; artificial neural network; genetic algorithm.

I. INTRODUCTION

Graphology is a research tool which can identify people's personality, intelligence, health and work capacity through analyzing their handwriting. It has been widely applied to many areas such as personality prediction, personnel selection, illness monitoring and court handwriting examination [1-2]. However, authoritative and scientific research about this subject is rather limited despite popular interest [3-4]. Especially in the study of the relationship between personality and handwriting, the psychometric reliability and validity have often been questioned and recent researchers infer that the correlations are marginal [3-7]. Throughout the studies of the last few decades, the methods used to analyze the relationship between personality and handwriting are limited to correlation analysis, regression analysis and some linear skills [3-4, 6]. Actually, the relationship may be rather complicated instead of simply linear. Conclusions based on these linear methods might be unconvincing. Therefore, this paper will employ artificial neural network to identify the association and examine the validity of handwriting analysis.

In our previous work [9], we presented a systematic study on the relationship between personality and handwriting of Chinese characters, through improving some methods of previous similar studies on western letters. This paper will not describe samples acquisition and feature extraction in detail, but will focus on the relationship between character spacing of Chinese characters handwriting and personality factors from Cattell's 16PF using artificial neural network. In the following sections, we will first discuss extraction of handwriting features and personality factors. Then we will introduce the neural networks adopted here along with their training methods. Finally, a detailed analysis of the network results will be presented. The whole process can be shown as follows:

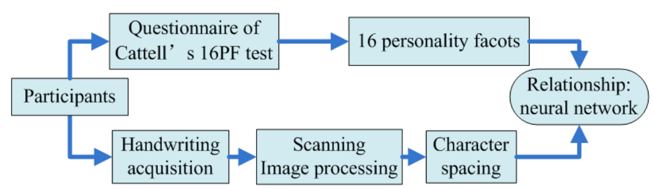

Figure 1. The chart of the whole process.

II. HANDWRITING FEATURES AND PERSONALITY FACTORS

A. Samples acquisition

One hundred nighty-two undergraduates consisting of 150 males and 42 females from Huazhong University of Science and Technology took part in the samples acquisition. The participants were asked to answer the questionnaire of Cattell's 16PF test one time, and to complete the form of handwriting acquisition three times as requested. The questionnaire of Cattell's 16PF test used here included 187 questions and some

978-1-4244-4994-1/09/$25.00 ©2009 IEEE

instructions. The questionnaire was used to acquire data on 16 personality factors of participants, namely "warmth, reasoning, emotional stability, dominance, liveliness, rule consciousness, social boldness, sensitivity, vigilance, abstractedness, privateness, apprehension, openness to change, self-reliance, perfectionism, tension" [5]. The acquisition of handwriting samples was from script written on white paper. The paper was comprised of two parts, one part was used to collect individual Chinese characters, and the other was used to collect connected Chinese characters. The part of connected characters was a poem by Li Po. All the scripts were recorded using a gel ink pen with 0.5mm refill (type: Cheng Guang M&G in China). The forms were then scanned into computer images by the MICROTEK scanner (type: MRS -9600U2) at a resolution of 400 dpi.

In view that the handwriting feature studied here is character spacing, this paper only considers the part with the connected characters. In order to present the results of neural networks concisely, only the samples of 150 male participants have been considered in this paper. Fig. 2 shows parts of one participant's script.

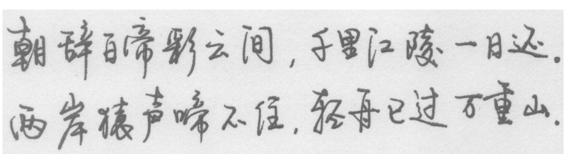

Figure 2. Parts of one participant's script.

B. Image processing

The image processing converts an entire image into many statistical features for later analysis. The process consists of three major steps: digitization, pre-processing and feature extraction, which are briefly described as follows.

Digitization: Each sample of handwriting acquisition is scanned into a digital image at a resolution of 400 dpi. Images are saved as 8 bits gray-scale TIF files.

Pre-processing: Pre-processing consists of such operators as binarization, noise removal, image skew correction, and segmentation, etc. In these operators, segmentation of connected characters is the most important and difficult one. There have already been many techniques used to segment connected characters, such as Viterbi algorithm, genetic algorithm, and methods based on histograms or recognition [11]. Each of these approaches has its own advantages and disadvantages in segmentation effect and time. In view that most of the samples only have touching strokes without overlapping or crossed strokes, this study first uses histograms of vertical projection to divide connected characters into several rough fields. Next we determine whether or not the text in the corresponding field is a single character through average character width. Finally, a flexible frame is introduced to merge and divide non-single character fields. Fig. 3 shows the segmentation results of one sample.

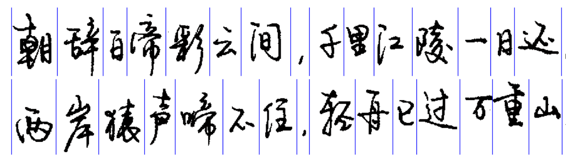

Figure 3. Segmentation results of one sample.

Feature extraction: Character spacing is an obvious feature for writers. This paper selects it as the subject to examine the relationship between handwriting and personality. Because the comma and period have great difference with Chinese characters, they are not taken into account here but are eliminated. The average character spacing and its normalized value can be expressed below:

$$cs = \frac{1}{24}\sum_{i=1}^{4}\sum_{j=1}^{6}\left[xl(i,j+1) - xr(i,j)\right] \tag{1}$$

$$\overline{cs} = \sum_{i=1}^{4}\sum_{j=1}^{6}\left[xl(i,j+1) - xr(i,j)\right] \Big/ \sum_{i=1}^{4}\left[xr(i,7) - xl(i,1)\right] \tag{2}$$

where $xl(i,j)$, $xr(i,j)$ refers to the left border and right border of the j th character in the i th sentence, as shown in Fig. 3; $\sum_{i=1}^{4}\left[xr(i,7) - xl(i,1)\right]$ is the total width of four sentences; cs is average character spacing and $\overline{cs}$ is normalized average character spacing. $\overline{cs}$ is the feature needed and studied in the analysis.

III. EXPERIMENTS OF NEURAL NETWORK

An artificial neural network is a computational model and mathematical model whose architecture essentially mimics the knowledge acquisition and organizational skills of the human brain. It has been demonstrated in theory that a neural network with one or more hidden layers can be trained to approximate any continuous function. This paper adopts three-layer feedforward network to examine the relationship between personality and handwriting. The input data is 16 personality factors of 150 male participants, so the number of the input layer's nodes is 16. And the output data is character spacing, so the number of the output layer's neurons is 1. The hidden layer uses log-sigmoid transfer function while the output layer uses linear transfer function. This study employs two network architectures with a different number of hidden layer's neurons and different training methods to complete two tasks.

Firstly, the 16-8-1 network is used to examine the strength of the association between personality factors and character spacing. And Levenberg-Marquardt (LM) algorithm is used to train the network.

Secondly, the 16-1-1 network is adopted to obtain brief relationship between the two kinds of variables. In the training process, it is found that the Levenberg-Marquardt algorithm will be caught in local optimal and the network cannot get well trained. Therefore, an alternate optimizing strategy of genetic algorithm (GA) and Levenberg Marquardt algorithm is presented. This strategy can be described by Fig. 4 and introduced as follows.

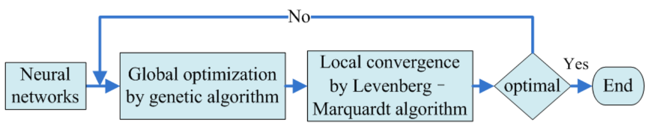

Figure 4. The flow chart of an alternate optimizing strategy.

- Step 1： Initialize parameters of neural network and genetic algorithm. Generate initial population for neural network. Every chromosome of the population is coded as a real number vector, which represents weights and biases of network.
- Step 2: Use genetic algorithm to evolve the population and search for the global optimum. The genetic algorithm here adopts mean square error as fitness function, and uses linear crossover operator and multi-point mutation operator. When the evolutional process arrives at a certain depth, proceed to step 3.
- Step 3: Use Levenberg-Marquardt algorithm to train the corresponding networks of different chromosomes, and to make parameters of these networks converge quickly. Because these networks often fall in local optimum, turn to step 2 to jump out of local values if the expected optimum hasn't been reached. Upon arriving at expectation, end the program.

As shown above, GA has two functions. One is searching for the global optimum and the other is jumping out of local optimums obtained by LM algorithm. Experiments show that the alternately evolving strategy can repeatedly exert the global optimizing ability of GA and the local convergent ability of LM algorithm, and have better convergent performance than single LM algorithm.

IV. RESULTS AND ANALYSIS

A. Results of 16-8-1 network

Of the 150 male participants, personality factors and character spacing of 120 ones are selected as the training data, and the other 30 are used for testing. The following two criteria are designed to describe the results of training and testing.

$$S(i)=\frac{|y'(i)-y(i)|}{y_{std}}, \quad \overline{S}=\frac{1}{n}\sum_{j=1}^{n}S(i) \tag{3}$$

$$E(i)=\frac{|y'(i)-y(i)|}{y_{ed}}, \quad \overline{E}=\frac{1}{n}\sum_{j=1}^{n}E(i) \tag{4}$$

where $y(i)$, $y'(i)$ denotes the desired output and the network's output of the i th participant respectively, so $|y'(i)-y(i)|$ represents the output error of the network; y_{std} denotes the standard deviation of character spacing of 150 participants while y_{ed} denotes the extreme difference; the value of n is 120 and 30 for training and testing respectively.

Table 1 shows the training and testing results. The upper part gives the values of $\overline{S}$ and $\overline{E}$ when training and testing, and the part below shows the distribution of all the values of $S(i)$ and $E(i)$. As can be seen from the left part, the average output error of training is very low and the average output error of testing is also low. At the same time the right part shows that there are 83.3% of testing samples whose $S(i)$ are less than 0.3, and $E(i)$ of all the testing samples are less than 0.2. These indicate that the 16-8-1 network is well suitable for examine the relationship, and the association is significant between character spacing and 16 personality factors. Moreover, it can be inferred from another aspect that analyzing people's personality by their handwriting is possible.

Table 1. The training and testing results of the 16-8-1 network.

			Training		Testing		
$\overline{S}$			7.76e-5		0.1732		
$\overline{E}$			1.58e-5		0.0353		
Range	<0.01	<0.05	<0.1	<0.2	<0.3	<0.4	<0.5
$S(i)$	6.7%	26.7%	46.7%	70.0%	83.3%	90.0%	93.3%
$E(i)$	26.7%	83.3%	93.3%	100%	100%	100%	100%

B. Results of 16-1-1 network

To obtain brief relationship between the two kinds of variables, the data of 150 participants are all used to train the 16-1-1 network. Table 2 gives the training results, in which the upper part gives the values of $\overline{S}$ and $\overline{E}$, and the part below shows their distribution. From the left part, it can be seen that $\overline{S}$ and $\overline{E}$ are not low enough, but still within a reasonable range. The right part shows $E(i)$ of 90.7% of samples are less than 0.3 though the values of $S(i)$ are not perfect enough. These results indicate that 16-1-1 network can reflect some association, although it is not quite adequate for linking the two kinds of variables.

Table 2: The training results of the 16-8-1 network.

				Training			
$\overline{S}$				0.6586			
$\overline{E}$				0.1341			
Range	<0.01	<0.05	<0.1	<0.2	<0.3	<0.4	<0.5
$S(i)$	5.3%	12.0%	14.7%	27.3%	36.0%	42.7%	46.0%
$E(i)$	12.0%	33.3%	46.0%	72.7%	90.7%	96.0%	100%

In order to examine how greatly different personality factors impact on character spacing, table 3 shows the weights of different personality factors to character spacing in the 16-1-1 network. It can be found that the factors 2 and 8, that is, reasoning and sensitivity, have the greatest impact on character spacing. And the weights of other factors are much less than these two factors' weights. In respect that the hidden layer uses log-sigmoid transfer function and the output layer uses linear transfer function, the following formulas can be used to express the relationship between character spacing and reasoning, sensitivity, if the whole impact of other factors is approximately regarded as a constant.

$$y_1 = \frac{1}{1+e^{-\left(a_2 x_2 + a_8 x_8 + b_1 + c\right)}}, \quad y_2 = k_2 y_1 + b_2 \tag{5}$$

where a_2, a_8 refers to the weights of reasoning and sensitivity respectively; b_1, b_2 refers to the biases of the hidden neuron and the output neuron respectively; k_2 refers to the weight of the hidden neuron to the output neuron; c is the constant approximately representing the whole impact of other personality factors. x_2, x_8 denotes the reasoning and sensitivity of participants; y_1 denotes the output of the hidden layer; y_2 denotes the output of the output layer, that is, character spacing.

Table 3: The weights of different personality factors to character spacing.

Factors ID	1	2	3	4	5	6	7	8
Weights	-7.0	-30.4	-1.0	10.6	0.5	10.3	-8.7	22.6
Factors ID	9	10	11	12	13	14	15	16
Weights	15.1	-14.8	10.1	0.7	-5.1	6.1	-1.4	-11.9

The values of a_2, a_8, b_1, b_2, k_2 can be imported directly from the training results, and they are -30.385, 22.648, -32.111, 0.1502, -0.0681 respectively. But the value of c needs to be identified by non-linear fitting. Taking reasoning and sensitivity as the independent variables, character spacing as the induced variable and formula 5 as the fitting function, it is not difficult to obtain the value of c, which is -79.130. As a result, the whole association between character spacing and reasoning, sensitivity can be presented in formula 6. In addition, Fig. 5 gives space distribution of character spacing, reasoning and sensitivity, and their fitting surface by formula 6. As can be found from Fig. 5, points of these three variables are very dispersive and stochastic, and the fitting surface is only suited to a certain extent. These indicate that the 16-1-1 network has certain defects.

$$y_2 = \frac{-0.068}{1+e^{30.385x_2 - 22.648x_8 + 111.241}} + 0.1502 \tag{6}$$

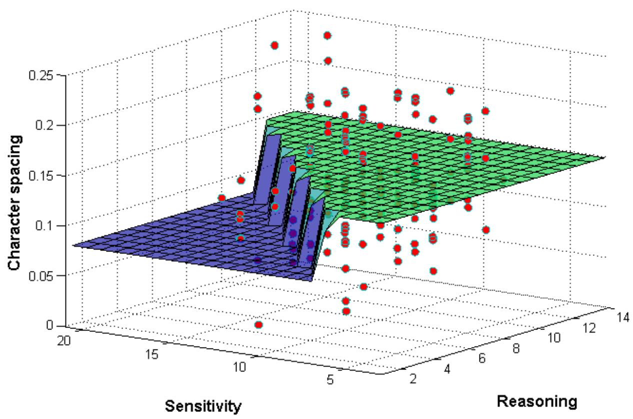

Figure 5. Space distribution of character spacing, reasoning and sensitivity, and their fitting surface by formula 6.

V. CONCLUSIONS

In this paper, artificial neural network is used to identify the relationship between personality and handwriting of Chinese characters. Results of 16-8-1 neural network indicate that the association is significant between 16 personality factors and character spacing. And from results of 16-1-1 network, it can be inferred that character spacing has strong correlation with reasoning and sensitivity.

However, the results of 16-1-1 network are not perfect. The training and fitting results both have some defects. This may be due to the following reasons. First, the number of the samples is not large enough, and samples have random error. Second, the relationship between handwriting and personality may not be deterministic but be a fuzzy relationship. The deterministic feedforward network may not be quite suitable here. And the fuzzy theory and fuzzy neural network may be more effective.

The value of this paper is making an attempt to adopt neural network to study the relationship between handwriting and personality, which was identified by linear methods in the previous work. Because there are not a set of complete features to represent the whole space of handwriting features, and Cattell's 16PF can reveal 16 complete personality traits which are manifested in people's behavior and habits, the paper calculates handwriting feature from personality factors. Results show that there is strong association indeed, which verifies the scientificalness and feasibility of graphology. But what is more useful in practice is inferring personality from handwriting, so a complete space of handwriting features is urgently needed.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the undergraduates from Huazhong University of Science and Technology for participating in the collection of handwriting and personality. And we express thanks to Shuaiying Wang for helping organize samples acquisition.

REFERENCES

[1] R.J Klimoski, A Rafaeli, Inferring personal qualities through handwriting analysis, Journal of Occupational Psychology, vol. 56, 1983， pp. 191-202.

[2] M.C Fairhurst, R.M Guest, C Ujam, Task-related population characteristics in handwriting analysis, Computer Vision, IET 2 (2) 2008， pp. 75-87.

[3] A Furnham, T Chamorro-Premuzic, I Callahan, Does Graphology Predict Personality and Intelligence?, Individual Difference Researchn, 2003, 1 (2).

[4] M.L Simner, R.D Goffin, A Position Statement by the International Graphonomics Society on the Use of Graphology in Personnel Selection Testing, International journal of testing 3 (4), 2003, pp. 353-364.

[5] http://www.16pfworld.com/keyinfo.html.

[6] M.R Shamsuddin, K.S Jazahanim, Z Ibrahim1, R.K.A.W Khan, A Mohamed, Graphology and Cattell's 16PF Traits Matrix (HoloCatT Matrix), in: Third 2008 International Conference on Convergence and Hybrid Information Technology, 2008, pp. 220-225.

[7] S Mutalib, SA Rahman, M Yusoff, A Mohamed, Personality Analysis Based On Letter 't'Using Back Propagation Neural Network, ICEEI 2007.

[8] Jun Sun, Yinghai Wang, Tianhong Pan, Genetic Neural Network and Its Application in Recognition of Tire Code, ICNSC 2008.

[9] Zhanghui Chen, Xiaohui Huang, Pengfei Chen, Wenlong Li, Siyao Zhu, Kangrui Zhou, Relationship between personality and handwriting of Chinese characters based on image processing, in press.

[10] S Mutalib, R Ramli, S.A Rahman, M Yusoff, A Mohamed, Towards Emotional Control Recognition through Handwriting Using Fuzzy, Information Technology, vol. 2, 2008, pp. 1-5.

[11] G.Louloudis, B.Gatos, I.Pratikakis, C.Halatsis, Text line and word segmentation of handwritten documents, Pattern Recognition, 2009.