

HANDWRITING RECOGNITION AND EVALUATION USING MACHINE LEARNING

Al Ahmed Faiyaz¹, Muhammed Sinan², Jaganathan S³

¹BE Computer Science And Engineering, Bannari Amman Institute of Technology, Sathy

²BE Computer Science And Engineering, Bannari Amman Institute of Technology, Sathy

³BE Computer Science And Engineering, Bannari Amman Institute of Technology, Sathy

ABSTRACT

Handwriting recognition for each document by its writer is a challenging problem due to huge variation in individual writing styles. The traditional approach to solving this would be to extract language dependent features like curvature of different letters, spacing between letters etc. and then we use a classifier like SVM to convert the document. In this project, it is to demonstrate deep learning based approach in identifying these features. We will pass small patches of handwritten images to a CNN and train it with classification loss. After the digital conversion of letters, it is then evaluated according to the necessities involved. In our project, we consider the examination answer sheets where they are manually corrected by the professors which take lot of time. Hence, to automate this process and to decrease the burden for the professors, we implement this evaluation system.

Keywords : Handwriting recognition,SVM,deep learning,CNN,answer sheets,professors

I. INTRODUCTION :

Handwriting also termed as brain-writing is a useful measure in identifying the characteristic personality traits of an individual. Handwriting analysts also known as graphologists can examine an individual's handwriting to predict the personality traits of the writer. Automated handwriting analysis can be used to examine personal traits of candidates during interviews accurately as the accuracy of an analyst highly depends on his skill set. The proposed tool will compliment the graphologists to increase their speed and efficiency in the analysis process. Machine learning approach like KNN with incremental learning, will be implemented to increase the efficiency of the tool.

II. LITERATURE SURVEY

The study of handwriting is quite an old concept tracing back to the seventeenth century. The first book to document these methods was written almost 400 years ago by Camilo Baldi. Known as the

father of graphology, Camilo Baldi, who was an Italian doctor of medicine and philosophy, performed systematic observations on handwriting samples in the year 1622. Since then, very detailed and extensive studies have been performed in this field. There are more than 2200 documented studies of handwriting analysis till date. [2]

A study conducted by the American Psychological Association's annual convention acknowledges that the use of computer technology in the field of handwriting analysis can be considered as a reliable tool for determining various traits like honesty, emotional stability, substance abuse risk and judgment.

This paper focuses only at the prime features of a page of handwritten sample that are page margins, line spacing, line direction, slant and zone ratios. To do this, the methodology used is scanning, preprocessing, feature extraction, analysis and finally, trait determination. This system was mainly designed to prove the validity of the graphology rules that were applied in the implementation of the system. This paper restricts its scope to macro analysis of the handwriting sample. There are no micro features like alphabet, loops etc. taken into consideration.

A paper based on artificial neural network explores the implementation of a machine learning approach in the field of handwriting analysis. This paper proposed a method to predict the personality traits of a person by analyzing the baseline, pen pressure and the letter “t” as found in the individual’s handwriting sample. These extracted features are then given as an input to the artificial neural network which in turn gives output as personality trait to the user. The future work discussed in this paper are, including more features of the handwriting like the size of the letters and the margins as inputs for personality trait determination to improve the system output[4].

The various algorithms and techniques used for the analysis have been discussed below:

2.1 Polygonalization :

Polygonalization is a method of subdividing the plane into polygons. This is the main technique used to find the slant of the baseline. In this technique of Polygonalization, a closed polygon is drawn around a single line of the scanned handwriting sample. The slope of the polygon in Fig. 1 is found using the coordinates of the polygon. This slope corresponds to the slope of the baseline. [4]

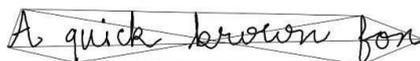


Fig. 1: Polygonalization

2.2 Template Matching :

Template matching is a technique in image processing used for finding small parts of an image which match a given image that is template image. From the scanned handwritten sample, individual lines are isolated and individual characters are identified. Finally, the identified characters are compared with all the template images by correlation and a match is found. [6]

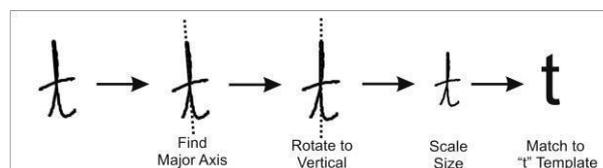


Fig. 2: Template Matching

III. PARAMETERS FOR ANALYSING HANDWRITING :

Graphologists are the handwriting analysts who identify the characteristics traits of a person merely by examining the handwritten samples of the individual. As this is a manual process, the skillfulness of the examiner defines the accuracy. However this is a time consuming and costly affair solely because of the human intervention. Thus, in the proposed methodology the focus would be on the development of a computer aided tool with minimum human intervention that would be able to predict the characteristic traits of a person intelligently. Baseline, writing pressure, spacing between letters, words and lines, size of letters, strokes connecting the letters, width of margins, starting strokes, ending-strokes, slant of word etc. are the most common parameters that help in identifying the personality traits of an individual through handwriting analysis. This paper would be focusing on the following four parameters: Baseline, Letter-slant, Height of the T-bar and Width of Margins.

3.1 Baseline:

The baseline is the feature which reveals a great deal of information as far as the personality of the writer is concerned. Baseline in a person's handwriting is the imaginary line along which the writer aligns the bottoms of the middle zone letters, when asked to write on a blank paper. Slanting downwards, rising upwards and level are the three most common baselines found in handwriting [4]. The technique that would be used to determine baseline is polygonolization. Different personality traits are associated with each of the above types of baselines. These characteristic traits are mentioned in the table below.

<i>Baselines</i>	<i>Corresponding traits</i>
Ascending(rising upwards)	Optimistic
Descending(slanting downwards)	Pessimistic
Level(Straight)	Balanced

Table 1: Baseline Characteristics

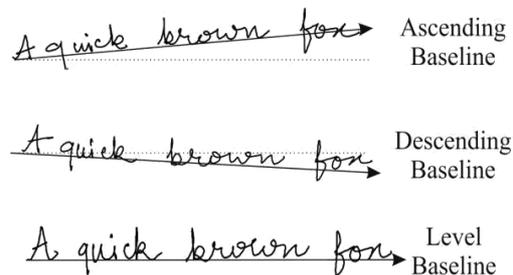


Fig. 4: Types of Baseline

.3.2 Letter Slant :

The slant of the letter is used to understand whether an individual's handwriting is inclined towards the right or towards the left or is it vertical. It has been observed that, around 77 percent of individuals write with a right slant, 15 percent with left slant and remaining 8 percent write vertically [7]. The slant of an handwriting is in relation to an individual's emotional direction and degree of sentimental control. The writer's connection between the inner and outer world is indicated by means of the slant of the letters.

Table 2: Slant Characteristics

Slant	Corresponding trait
Extreme left	Fear of the future, defensive, early rejection.
Extreme right	Lack of self-control, impulsive, unrestrained, intense, very expressive, low frustration tolerance.
Left	Reflective, independent, not sympathetic, difficulty in adapting and expressing emotions.
Right	Expressive, confidence in convictions, freedom of thought, extrovert, future orientation.
Vertical	Head controls over heart, independent emotional nature, and works independently.

Table 3: 't' Bar Characteristics

<i>Position of t-bar</i>	<i>Corresponding trait</i>
Crossed high (but above stem)	High self-esteem: This reveals confidence, ambition, and the ability to plan ahead, high goals, high personal expectations and an overall good self-image.
Crossed above middle zone	Moderate self-esteem: It indicates a practical and successful person.
Crossed low on stem	Low self-esteem: This person fears failure and resists change. He is rarely successful enough in his own eyes despite his accomplishments and resists change.
Crossed above the stem	Dreamer: This person's goals and dreams lose touch with reality.

Table 3: 't' Bar Characteristics

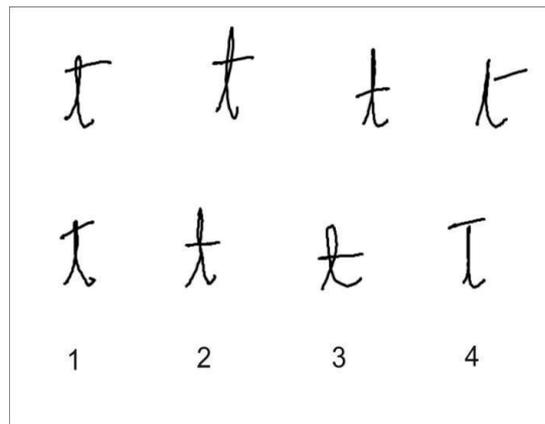


Fig. 5: Types of 't' Bar

3.4 Margin :

The margin is generally considered to define the layout of the page. Any individual assumes a particular margin while writing on a blank paper. The blank spaces on the left, right, top and bottom on a page comprises the margin. In this paper, different types of left and right margins namely- wide right, wide left, no margin at all and evenly spaced margin will be considered. Following are the characteristic personality traits of individuals associated with the above mentioned types of margin [1].

Margin Orientation	Corresponding trait
Wide left	Courageous.
Wide right	Avoids future and a reserved person.
No margin	Insecure and devotes oneself completely.
Even margin	Self disciplined and balanced.

Table 4: Margin Characteristics

IV. PROPOSED APPROACH

This work aims at obtaining personal characteristic traits of individuals especially in the age group of 20-35 years when they face many interviews.

4.1 Generation of Training Data set

Sample Training set will be generated using 100 samples of handwriting, which will be examined by a professional. Graphologist. Using Image Processing tools the feature vectors matrix of the same handwriting samples will be generated. The Feature Vector Matrix is the mathematical representation of the handwritten text in form of feature vectors. Based on the characteristics corresponding to particular trait, as identified by the graphologist, classes will be created. Initially, each FV Matrix, created for a particular handwriting sample, will be manually mapped into its corresponding class as suggested by the graphologist [6]. In this way a data-set will be generated that contains individual classes corresponding to each trait. Later a supervised machine learning algorithm would be used for mapping new samples into previously identified classes.

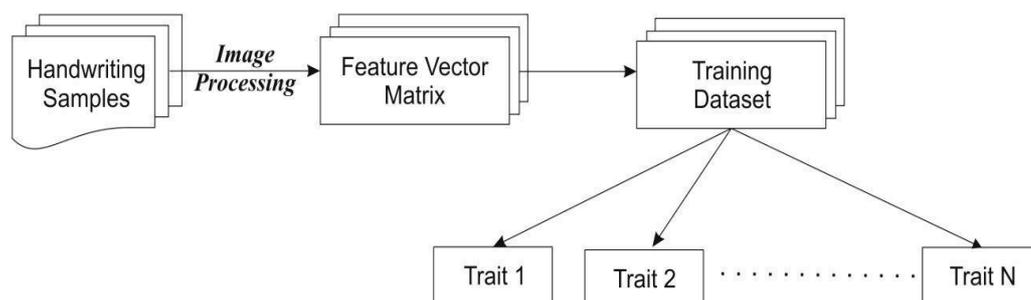


Fig. 6: Process Flow Diagram Generation of Training Data-set

4.2 Trait Identification Based on Trained Data

Once the dataset is trained, it will be easy to identify the traits corresponding to a new sample of handwriting. The Feature Vector Matrix for a new handwriting sample will be created and then its similarity with the trained dataset would be calculated using the similarity matrix method. K-NN.

classifier will be used to identify the class which is most appropriate for the handwriting sample, based on the similarity matrix. In this way the handwriting would be identified in a class that will map the handwriting to its corresponding trait. After the results are produced, it would be stored in trained dataset. By storing the new sample into the class to which it is mapped will be able to implement the incremental machine learning, increasing the efficiency of future results.

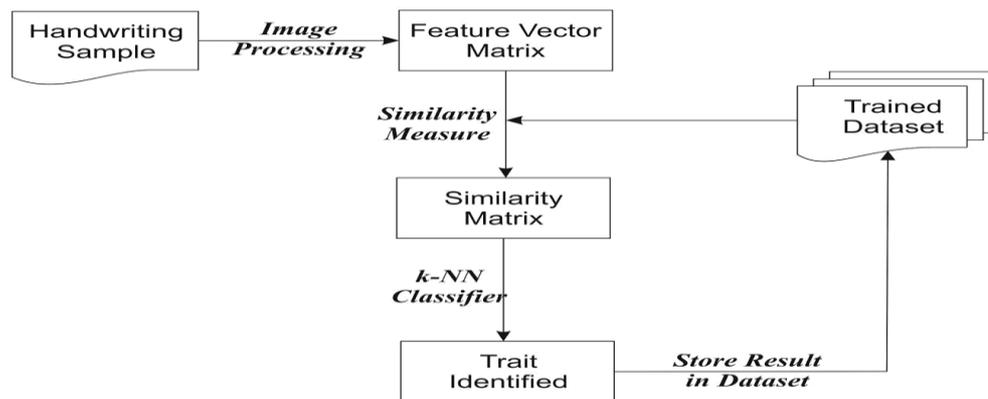


Fig. 7: Process Flow Diagram for Classification of Traits

V. CONCLUSION :

This paper has proposed a methodological analysis to predict the accurate personality traits of an individual from the features extracted from handwriting using a machine learning approach. This paper explores the personality traits revealed by baseline, margin, slant of the words and height of t-bar of a person's handwriting. These features will be extracted from the handwriting samples into feature vectors which would be compared with an initially trained data set; and then mapped to the class with corresponding personality trait. The baseline would be evaluated using the method of Polygonolization while margin will be calculated using the method of vertical scanning. The height of the t-bar on the stem of the alphabet 't' and word-slant would be calculated using template matching.

The proposed system can be used as a complementary tool by the graphologist to improve the accuracy of handwriting analysis and also make the process fast. The future work can include more features from the micro approach of handwriting analysis like the loops of alphabet 'f' and 'l', gradient, concavity of letters and so on in order to predict more accurate results.

REFERENCES

- [1] Champa,H.N.andAnandakumar, K.R,“Automated human behavior prediction through handwriting analysis”, Integrated Intelligent Computing (ICIIC),
- [2] Artificial Neural Network ,Available at: <http://cogsci.stackexchange.com/questions/8509/dynamic-al-systems-theory-as-a-metaphor-in-psychology-is-it-useful-or-not> Spector, A. Z. 1989. Achieving application requirements. In Distributed Systems, S. Mullender

[3] 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.

[4] Sofianita Mutalib, Shuzlina Abdul Rahman, Marina Yusoff1, Azlinah Mohamed
“Personality Analysis Based On Letter „t“ Using Back Propagation Neural Network”, Proceedings of the International Conference on Electrical Engineering and Informatics Institute Teknologi Bandung, Indonesia June 17-19, 2007.