Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 165 (2019) 384–390

Procedia Computer Science

www.elsevier.com/locate/procedia

INTERNATIONAL CONFERENCE ON RECENT TRENDS IN ADVANCED COMPUTING 2019, ICRTAC 2019

Handwriting Analysis based on Histogram of Oriented Gradient for Predicting Personality traits using SVM

Aditya Chitlangia[a], G.Malathi[a]

[a]*Vellore Institute of Technology, Chennai, India*

Abstract

Handwriting Analysis is a method to understand and predict the characteristic traits of a person based on his handwriting style. Graphology is the scientific term used for handwriting analysis. Professional handwriting examiners, called graphologists, manually study and understand the handwriting of an individual to classify the writers personality. Nevertheless, the manual process of handwriting analysis is time-consuming, costly and depends majorly on the skills of the graphologists. To make this process computerized we extracted several features of handwriting samples and classified the writer into 5 personality traits namely Energetic, Extrovert, Introvert, Sloppy and Optimistic. Histogram of oriented gradient(HOG) extracts the features from the handwriting sample of the writer which serves as an input for the Support Vector Machine model to give output as the personality trait of the person. For this paper, digital handwriting sample data of 50 different users were collected. The proposed system predicts the personality trait of a person with 80% correctness using the Polynomial kernel. In this paper, we propose a computerized method for personality trait prediction based on the users handwriting. Two different methods are applied to the same handwriting sample data to measure and compare the performance of the proposed system.

Keywords: Graphology; Support Vector Machine; Histogram of oriented Gradient; Personality Trait; Behavior Analysis

*Corresponding author. Aditya ChitlangiaTel:8003893300

E-mail address: <aditya.chitlangia2016@vitstudent.ac.in

1877-0509
10.1016/j.procs.2020.01.034

1. Introduction

Graphology is the scientific methodology to identify, evaluate and understand an individual's temperament through his handwriting. Strokes, spaces, patterns, and pressure applied while writing can predict specific personality traits [1]. Handwriting determines true personality like honesty, fears, behavior among many others. Handwriting examiners called graphologists analyze many characteristics of handwriting samples to detect the writer's character traits. There are many techniques to predict the personality traits of an individual. The process starts with the collection of handwriting samples of various individuals on a plain white paper. Handwriting analysis needs several preprocessing steps such as resizing, noise removal and binarization, etc. on sample data.

The proposed method automates the technique of graphology. Compared with other procedures, such as performing a manual examination of handwriting samples of the user; proposed computerized handwriting analysis is much faster, precise and economical. The rule-based approach [2] also plays a significant role in graphology. An individual can conveniently use digital handwriting data as input to the computer and the automated algorithm calculates the features using image processing and machine learning techniques predicting the writer's personality trait.

This research focuses mainly on the extraction of features from digital samples and classifying the personality of user based on it. All feature extraction and calculation are computerized making this technique simple, optimized and easy to implement.

2. Proposed Methodology

The graphologist analyzes the handwriting of a person written by them on a piece of paper which is very timeconsuming. The accuracy of the manual graphology depends totally on the experience and the knowledge of the graphologist.Several works have been proposed in this field [3,4,5] which follows 4 mains steps: Appropriate data collection, Pre- processing of the data, extracting features from the data and Classifying the personality of the user. The proposed system follows the approach as shown in Fig. 1

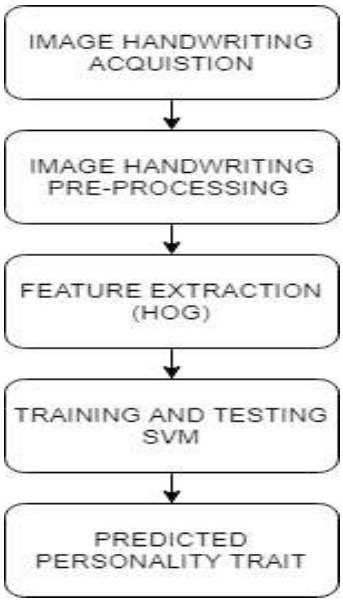

Fig. 1. Work Flow of the proposed system

2.1. Collection of Handwriting samples

Handwritten samples of various people are collected and used for this experiment. 50 different writers are asked to write several texts with the same content. All the handwritings are cursive. The sample handwriting is written on a plain page with no margin as given in Fig 2.

Fig. 2. Handwriting dataset of different individuals with labelled personality traits

2.2. Pre-Processing of Handwriting Images

In the pre-processing stage, noise removal from the handwritten image sample is done using adaptive thresholding followed by resizing the handwriting image sample to correctly orient the image. In pre-processing stage removal of unwanted characters (dots, scribbling, etc.) and smoothing of digital handwriting sample is done.

2.3. Feature Extraction

The most crucial step to attaining high recognition accuracy is the selection for the best method of feature extraction. The feature extraction technique is used for reducing the dimensionality of the input sample data. The reduced output is represented as a feature vector. Feature extraction methods differ for different applications. Techniques that are successful for one application, may get failed for other application [6]. The features extracted must possess the important characteristics of the letter differentiating it from others. Therefore, in this paper, Histogram of oriented Gradient(HOG) technique is adopted for extracting the features from handwritten images.

2.3.1. *Histogram of Oriented Gradient*

Histogram of Oriented Gradient(HOG) [7] technique was proposed by N. Dalal and B. Triggs for application in the detection of the human body. Today, this technique is the most popular and successful descriptor in image processing and computer vision for object detection [8]. HOG works by counting the occurrences of gradient orientation in a localized portion of the sample images. The important idea behind HOG descriptors is that local object appearances and shape within an image can be described by edge directions or distribution of intensity gradients. HOG technique coverts the digital handwriting sample into square grids. After that according to the central difference, the edge or histogram of gradient direction is computed. The calculation of HOG features is done by taking orientation histograms of edge intensity in a local region[9]. In this paper, features are extracted using HOG from all locations of the grid on the handwritten image as a candidate of the feature vector.

3. Classification

A classifier is used to predict the personality by extracting features from the handwriting sample using HOG [12]. These extracted features are calculated and stored as models and further used in trained classes. The common method used in the classification stage is the use of Support Vector Machines [11] as given in Fig 3. It is used to classify the personality of the user among the 5 labeled personalities namely Energetic, Extrovert, Introvert, Sloppy and Optimistic via the features captured by HOG.

3.1. *Support Vector Machine*

The classifier used here is the Support Vector Machine(SVM). With better accuracy and time efficiency, SVM works better than Neural networks in this situation. In a Multiclass SVM, different labels are assigned to instances using SVM where different classes are assigned to elements for distinguishing them. Support Vector Machine works by taking x as train data from some space R^D where, x={x^1, x^2, x^3,...x^n} such that x belongs to R^D. We give their class label y={y^1,y^2,y^3,...y^n} as to which the x^i belongs to.

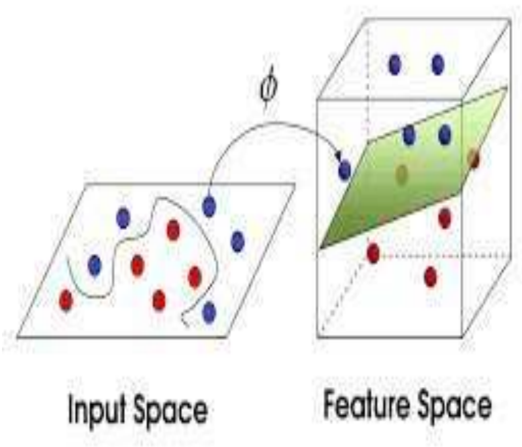

Fig. 3. Principle of Support Vector Machine [11]

The polynomial kernel has been used by us since it gives better results in comparison to nonlinear kernels like sigmoid and RBF on non-linear features. The Polynomial kernel maps the input space to a higher dimension in a non-linear fashion. The polynomial kernel is outlined asequation(1)

$$K(x,y) = (x^T y + c)^d \tag{1}$$

Where, x and y are vectors of features extracted from training the digital handwriting samples in the input space, and c≥ 0 is a trading off parameter.

Here in kernel, K is the inner product in a feature space based on some mapping Φ : as given by equation (2)

$$K(x,y) =< \Phi(x),\Phi(y) > \tag{2}$$

The classification method follows 3 steps:

1. Creation of the vectors from input features.
2. Applying the Polynomial kernel to map the feature space into a higher dimension.
3. Computation of a hyper-plane which separates the feature space to different classes of sample vector.

4. Experiment

The experimentation is conducted in 2 stages, A and B. Stage A takes 90% of sample data as training data and 10% as testing data. While stage B comprises takes one data as test data and remaining as training data.At step one, every writers handwriting sample was stored in electronic type. Subsequently, the manual evaluation of every individual was achieved to categorize their characteristics to among the cited Energetic, Extrovert, Introvert, Sloppy, Optimistic. 90%of its particular own consequences along with this census statistics sample have been utilized to determine the user's hand-writing information while 10% samples have been used as testing data. It ends at 80% of precision. While at stage B tested with a single sample while the evaluation sample although we now make utilize of exactly precisely the way to prepare the machine. Inside this instance, the precision stays more or less the same.

Fig. 4 shows the relation of different handwriting styles with characteristic traits and behavior of a user based on his style of writing the letter t given by Champa H. N. and K. R. Ananda Kumar [10].

S.No.	Writing Categories	Psychological Personality Behavior
1.	Large Letters	Likes being noticed, stands out in a crowd
2.	Small Letters	Introspective, not seeking attention, modest
3.	Medium Letters	Adaptable, fits into a crowd, practical, balanced
4.	Right Slant	Sociable, responsive, interested in others, friendly
5.	Left Slant	Reserved, observant, self-reliant, non-intrusive
6.	Vertical	Practical, independent, controlled, self sufficient
7.	Light Pen Pressure	Strong emotion, successful, emotion last for long time
8.	Heavy Pen Pressure	Try to avoid energy draining situations
9.	Far Spacing Letters	Openness of sentiment and intelligence
10.	Close Spacing Letters	Closeness of sentiment and intelligence
11.	Far Spacing Words	"give me breathing space"
12.	Close Spacing Words	Wish to be with others, intrusive
13.	Raising Baseline	Optimistic, upbeat, positive attitude, ambitious, hopeful
14.	Falling Baseline	Tired, overwhelmed, pessimistic, not hopeful
15.	Straight Baseline	Determined, stays on track, self motivated, controls emotions, reliable, steady

Fig. 4. Characteristic traits and behavior based on writing style of letter t

The accuracy rate of the proposed system for method A and method B with polynomial kernel is shown in Table 1

Table 1. Accuracy table

Kernel	A	B
Polynomial Kernel	80%	80%

5. Result Set

The experiment was conducted on a handwriting sample collected from 50 people. The dataset was separated into training, testing and Machine learning model was applied on it. In Fig. 5 a handwritten sample of a user is taken. Then Gaussian blur is applied on it to reduce image noise as shown in Fig. 6. After that adaptive thresholding is applied to separate desirable foreground image objects from the background based on the difference in pixel intensities of each region as show in Fig. 7. After that it is passed through the classification algorithm which gives correct prediction of Extrovert personality trait in Fig. 8.

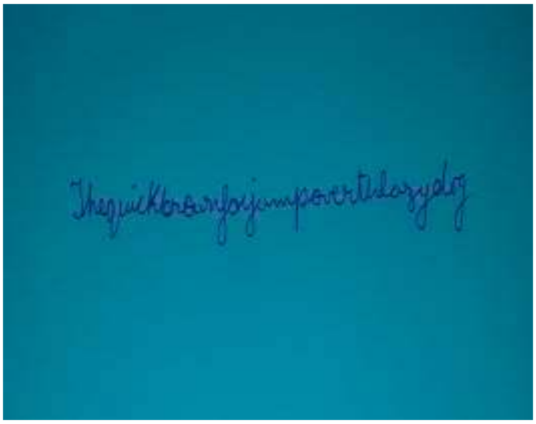

Fig.5.Inputhandwritingsample

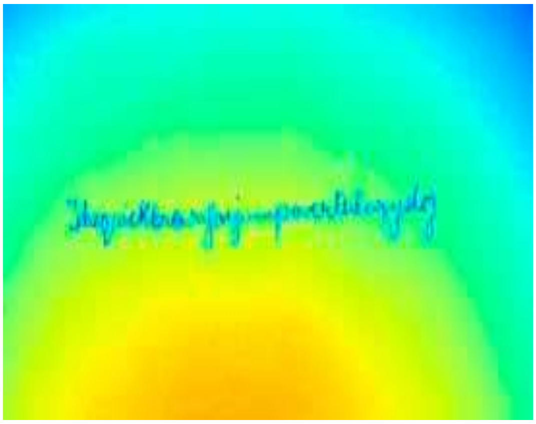

Fig.6.Gaussianblurtoreduce noise

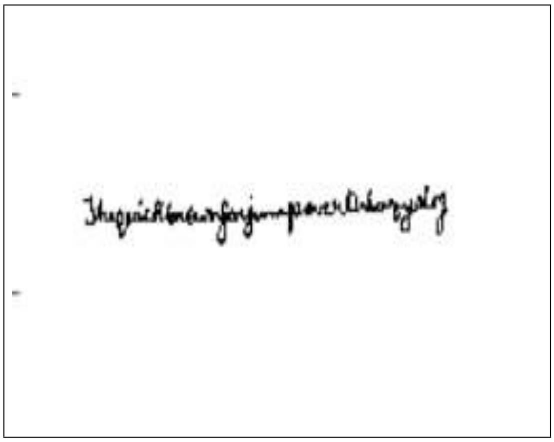

Fig.7.Adaptivethreshold

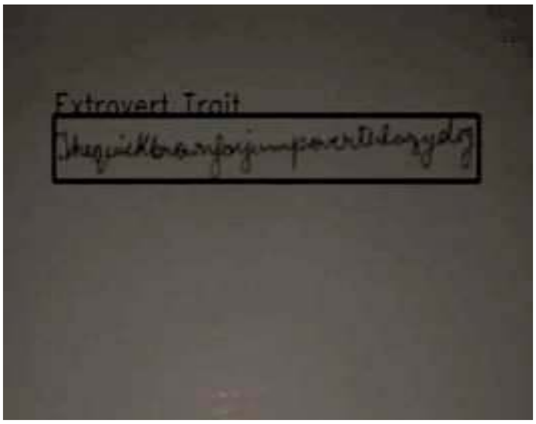

Fig.8.PredictedPersonalityTrait (Extrovert)

6. Conclusion

This paper proposes a computerized system for prediction of the characteristic traits of a human being based on their handwriting styles using Support Vector Machine classifier. SVM takes the features extracted using HOG technique as input and classifies the personality of the individual writer into one of 5 personality traits (Energetic, Extrovert, Introvert, Sloppy and Optimistic) as the output with 80% accuracy.

7. Future Work

Currently the data set is very limited, it only has 10 samples per class. It is best to have more than 500 samples per class to be able to detect the characteristics of the handwriting of the users more precisely.For now, the system only uses HOG for feature extraction to make the graph logical analysis, but in the future we will extract features of each character using the slant of words and letters, spaces between words and letters and pen pressure thus make the analysis richer and closer to reality, increasingly closer to one given by a professional.

References

[1] Rizvi, FarhanaKhiyal, Malik. (2017). Personality prediction Through Offline Handwriting Analysis. Journal of Multidisciplinary Engineering Science Studies. 3. 1282-1288.

[2] 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 Sais BioSciences Research Institute Pvt. Ltd., 2009

[3] S.H. Cha, S. Srihari, Multiple Feature Integration for Writer Verification, 7th International Workshop on Frontiers in Handwriting Recognition; IWFHR VII, Amsterdam, The Netherlands, pp 333-342, 2000.

[4] U.V. Marti, R. Messerli, H. Bunke, Writer Identification Using Text Line Based Features, Proc. ICDAR01, Seattle (USA), pp 101-105, 2001

[5] A. Srihari, S. Cha, H. Arora, S. Lee, Individuality of Handwriting: A Validity Study, Proc. ICDAR01, Seattle (USA), pp 106-109, 2001.

[6] A. Lawgali, A Survey on Arabic Character Recognition, International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, no. 2, (2015), pp. 401426.

[7] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection, CVPR, (2005)

[8] R. Ebrahimzadeh and M. Jampour, Efficient Handwritten Digit Recognition based on Histogram of Oriented Gradients and SVM, International Journal of Computer Applications, vol. 104, no. 9, (2014), pp. 1013.

[9] T. Kobayashi, A. Hidaka and T. Kurita, Selection of Histograms of Oriented Gradients Features for Pedestrian Detection, (2008), pp. 598 607.

[10] Champa H. N., Dr. K. R. AnandaKumar, Artificial Neural Network for Human Bahavior Prediction through Handwriting Analysis, International Journal of Computer Applications (0975-8887) Volume 2 No.2, May 2010.

[11] Prasad, Shitala, et al. Handwriting Analysis Based on Segmentation Method for Prediction of Human Personality Using Support Vector Machine. International Journal of Computer Applications, vol. 8, no. 12, 2010, pp. 2529., doi:10.5120/1256-1758.

[12] Lawgali, Ahmed. (2016). Recognition of Handwritten Digits using Histogram of Oriented Gradients. International Journal of Advanced Research in Science, Engineering and Technology. 3.