RESEARCH ARTICLE OPEN ACCESS

SVM Based Identification of Psychological Personality Using Handwritten Text

Syeda Asra[1,] Dr.Shubhangi[2]
([1]Department of Computer Science & Engineering, Appa Institute of Engineering & Technology, Kalaburagi, India)
([2]Visvesvaraya Technological University, Regional Centre, Kalaburagi, India)

ABSTRACT
Identification of Personality is a complex process. To ease this process, a model is developed using cursive handwriting. Area based, width based and height based thresholds are set for only character selection, word selection and line selection. The rest is considered as noise. Followed by feature vector construction. Slope feature using slope calculation, shape features and edge detection done using Sobel filter and direction histogram is considered. Based on the direction of handwriting the analysis was done. Writing which rises to the right shows optimism and cheerfulness. Sagging to the right shows physical or mental weariness. The lines which are straight, reveals over-control to compensate for an inner fear of loss of control.The analysis was done using single line and multiple lines. Simple techniques have provided good results. The results using single line were 95% and multiple lines were 91%.The classification is done using SVM classifier.
Index Terms— Support Vector Machine (SVM), Personality, Human Behavior.

I. INTRODUCTION

Human behavior [1] refers to the range of behaviors exhibited by humans and which are influenced by culture, attitudes, emotions, values, ethics, authority, rapport, hypnosis, persuasion, coercion and genetics. Human behavior is experienced throughout an individual's entire lifetime. It includes the way they act based on different factors such as genetics, social norms, core faith, and attitude. Behavior is impacted by certain traits each individual has. The traits vary from person to person and can produce different actions or behavior from each person. Everyday conceptions of personality trait [2][3] make two key assumptions. First are stable over time. Most people accept that an individuals' behavior naturally varies somewhat from occasion to occasion, but would maintain the core consistency which defines individual's true nature like the unchangeable spots of the leopard.

II. RELATED WORK

A variety of systems have been designed and implemented. The previous and the most recent system consisted of single line and only slope features were considered based on line regression [4]. The network was trained using ANN [5] and SVM [6].The performances were compared using ANN and SVM. The results with SVM were found to be 97% and that of ANN was 85%.

III. PROPOSED METHODOLOGY

Trying to understand people's inner motivations is notoriously uncertain science. Personality identification offers one of the few routes into this world. Handwriting is brain writing.Handwriting Analysis or Graphology [7] is a scientific method of identifying, evaluating and understanding personality through the strokes and patterns revealed by handwriting. Accuracy of handwriting depends on how skilled the analyst is. Perhaps Human intervention in handwriting analysis is effective but prone to fatigue and inaccurate. Hence a system is developed to automate this process.

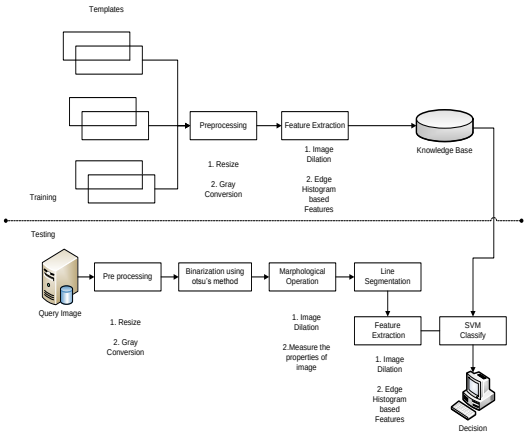

Fig 1: Block diagram of proposed system

3.1 Image Handwriting Acquisition and Database Creation

Data samples of 500 in number were collected from people belonging to different works of life both equally from males & females. In order to make the results time and mood invariant the samples were collected in different days at different points of time. A4 paper with black ball point pen with a hard surface was used for writing .Three paragraphs were given to write. The images were scanned using laser jet scanner with a resolution of 2528x3507 pixels and 300dpi.

3.2 Noise Removal & Image Handwriting Pre processing

A global threshold of the image is found using Otsu's method [8].The image is then converted into binary image as shown in figure 2. Morphological based operation, image dilation is performed to bridge the gap between the characters in the cursive handwriting as shown in figure 3.

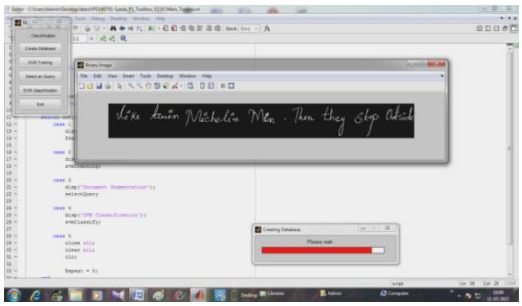

Figure 2: Gray image converted into binary image

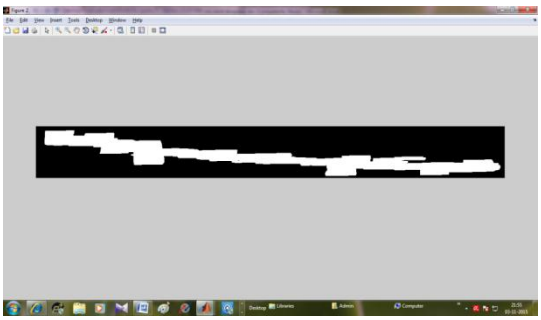

Figure3: Image Dilated

3.3 Feature Detection and Psychological Representation

Three features were extracted relevant for recognition, slope feature, shape feature and direction of histogram.

3.3.1 Slope Feature

Slope feature was used to find the slope of the line

3.3.2 Shape Feature

The shape feature was extracted using area, perimeter, form factor, major axis, minor axis, roundness compactness, density

3.3.3 Edge Orientation Histogram

The basic idea is to build a histogram [9] with the directions of the gradients of the edges (borders or contours). It is possible to detect edges in an image but it in this we are interested in the detection of the angles. This is possible through Sobel operators [10].

IV. CLASSIFICATION

4.1 SVM's Overview [11]

SVM classification is based on the idea of decision hyper planes that determine decision boundaries in input space or high dimensional feature space. SVM constructs linear functions (hyper planes either in input space or in feature space) from a set of labelled training dataset. This hyper plane will try to split the positive samples from the negative samples. The linear separator is commonly constructed with maximum distance from the hyper plane to the closest negative and positive samples. Intuitively, this causes correct classification for training data which is near, but not equal to the testing data. Throughout training phase SVM takes a data matrix as input data and labels each one of samples as either belonging to a given class (positive) or not (negative).SVM treats each sample in the matrix as a row in a input space or high dimensional feature space, where the number of attributes identifies the dimensionality of the space. SVM learning algorithm determines the best hyper plane which separates each positive and negative training sample. The trained SVM can be deployed to perform predictions about a test samples (new) in the class. Nonlinear problems in SVM are solved by mapping the *n*- dimensional input space into a high dimensional feature space. Finally in this high dimensional feature space a linear classifier is constructed which acts as nonlinear classifier in input space.

4.2 Radial Basis Function [12](RBF) and Support Vector Machines (SVM) networks

- The belief is that a classification problem cast into a higher dimensional space becomes more likely separable than in a lower dimensional space.
- RBF network can be used find a set weights for a curve fitting problem. The weights are in higher dimensional space than the original data.
- Learning is equivalent to finding a surface in high dimensional space that provides the best fit to training data.
- Hidden layers provide a set of functions that constitute an arbitrary basis for input patterns when they are expanded to the hidden space; these functions are called radial basis functions.
- Input layer + only one hidden layer responsible for spanning hidden space to which input vectors are mapped + a linear output layer.
- The functions are nonlinear. → In a pattern recognition task this often guarantees better separation.
- The functions map the samples from the input space to hidden space, in which the same samples are hopefully separable.

4.3 Radial Basis Function (RBF)

4.3.1 Gaussian functions

$\varphi(r) = \exp(-r^2/2\sigma^2)$ for some $\sigma > 0$ and $r \in R$ (Eqn. 1)

Gaussian functions are probably the most used. In general, the selection depends on the application.

- Support vectors are a set of training samples extracted by the algorithm used to find the optimal plane.
- The distance between the separating hyper plane and the closest data point is called the margin of separation (denoted ρ in the book).
- The goal of a support vector machine is to find the particular hyper plane that maximizes the margin of separation.
- Furthermore, the goal is to determine such a separating hyper plane which is optimal
- Optimality is defined here as $d_i \cdot (w^T x_i + b) \geq 1$ for i = 1, . . . N training samples, where $d_i = +1$ or $d_i = -1$.
- Also the weight vector w should minimize the cost function: $\Theta(w) = 1/2 w^T w$
- Obtaining the optimal solution
- The two optimality constraints can be expressed in a form of single cost function.

$$J(\mathbf{w}, b, \alpha) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_{i=1}^{N} \alpha_i [d_i(\mathbf{w}^T\mathbf{x}_i + b) - 1],$$

(Eqn. 1)

Where α_i are so called Lagrange multiplies.

- The optimal solution is found at the saddle point of J(·) (J is minimized with respect to w and b and maximized with respect to α)
- The saddle point is found by differentiating J first with respect to w and then b and setting the differentials equal to zero.
- Input data is transformed from input space with dimension m_0 to a higher dimension feature space with dimension m_1.
- The separating hyper plane i.e. the weights w_j is estimated in the feature space.
- The inner-product kernel can be used to construct the optimal hyper plane in the feature space without having to consider the feature space itself in explicit form.
- Inner product kernel loosely speaking: It is a function replacing activation function of a single neuron network**.**
- The inner product kernel is a function that is used to find the optimal hyper plane for a support vector machine network.
- Inner product kernel is a function:

$$K(x, x_i) = \varphi^T(x)\varphi(x_i), \quad \text{(Eqn. 3)}$$

Where φ is a set of transformation function $\varphi_j(\cdot)$, j = 1, . . . m_1. m_1 is the dimensionality of the higher order space (original dimension m_0). x_i is i^{th} training sample i.e. it is a known sample unlike x.

- Depending on how the kernel is defined different learning machines can be built using SVM. Such machines are polynomial learning machines, RBF networks, and two-layer perceptrons.

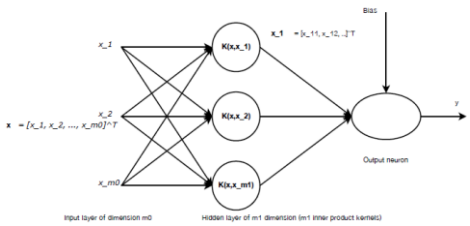

Figure 4: Architecture of Support Vector Machine

- x = [x 1, x 2 . . . m0] ^T is input sample.
- K (·, ·) are inner product kernels. I x 1, x 2 . . . x m1 are examples of input data. I after applying inner product kernels to input sample x, the data sample is mapped from input space to feature space.
- An optimal hyper plane is searched for in the feature space.
- Note the indexing: input vector element is denoted by x 1 whereas x 1 denotes training sample 1.

V. CO-RELATION BETWEEN FEATURE DETECTION AND PERSONALITY

5.1 Up-Hill Line(Cheerfulness)

Writing "uphill" reveals the applicant worthy of further trait-match evaluation. He's optimistic, ambitious and cheerful.Healthy mental energy & can stay busy active & constantly on the go.

5.2 Down –Hill Line (Weariness)

Writing "downhill" may be caused by a temporary depression, ill health, or physical fatigue. If this applicant has special skills necessary in your business, it would be wise to have him return on another day and obtain another sample of his writing. The first down slant sample could be just a temporary discouragement from job-hunting. If the second sample has the same down slanted appearance, it is a warning of an ingrained pessimism which keeps this applicant on the job-seeking circuit. Negative, disillusionment, constant disappointment.

5.3 Constant Line(Reveals over control to compensate for an inner fear)

The writer of a firm, straight, even baseline controls his moods, allowing him to go directly toward his goals without getting side-tracked. Reveals over control to compensate for an inner fear Healthy mental energy & can stay busy active & constantly on the go.

VI. RESULTS AND DISCUSSIONS

6.1 Analysis of Single Line in Handwritten Text

Line segmentation was done using morphological operations. Figure shows a binary image classified as Ideal. This indicates lines with constant slope. Figure shows the line classified as weariness. This indicates temporary depression and disillusionment and pessimism. The last classification is cheerfulness as shown in the figure. This indicates people with healthy mental energy and optimism.

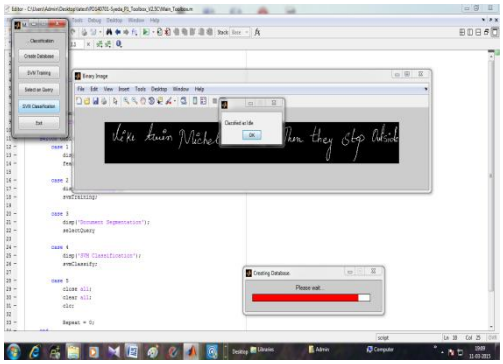

Figure. 5: Single line classified as ideal

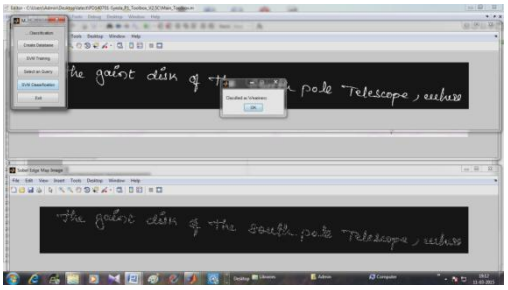

Fig 6: Line classified as weariness

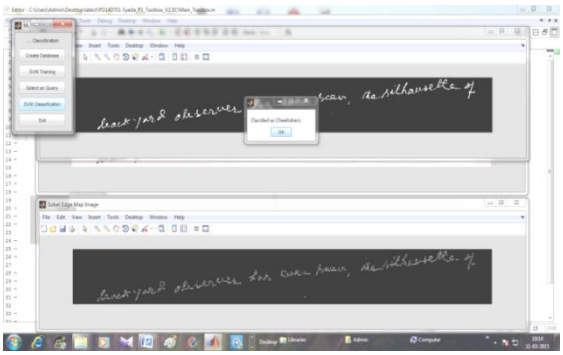

Figure 9: Line classified as cheerfulness

6.2 Analysis of Multiple Lines in Handwritten Text

The experimentation is carried for multiple lines. The SVM RBF kernel used has given outstanding results.

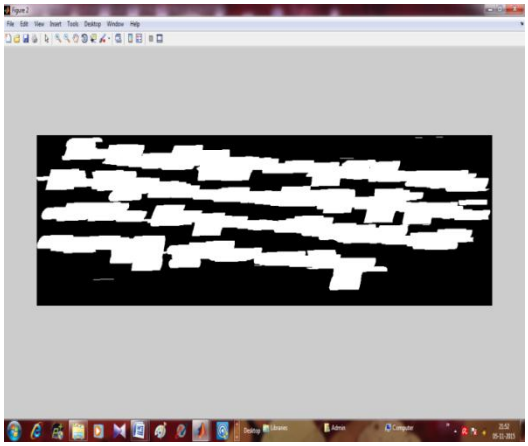

Fig. 8: Multiple lines considered for classification

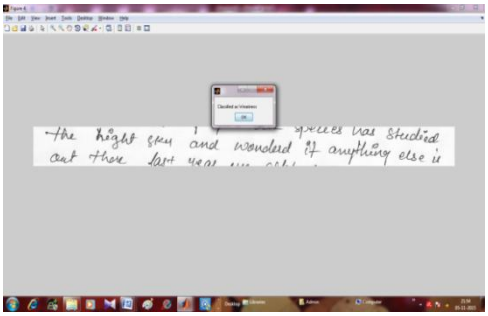

Figure 9: Multiple lines classified as weariness

7 Handwriting Analysis of Celebrities and Comparison with Graphologists

7.1 Hand Writing Analysis of Mr. Mittal (CEO of Arcelor Mittal)

The figure shows the handwritten text of famous personality Mr.Mittal. The network classified his handwriting as Ideal. The graphologists say he is controlled. Head rules his heart.

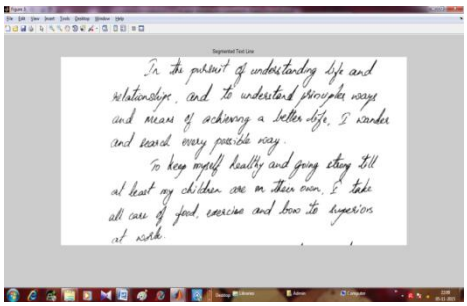

Figure10: Handwriting Sample of Mr. Laxmi Mittal

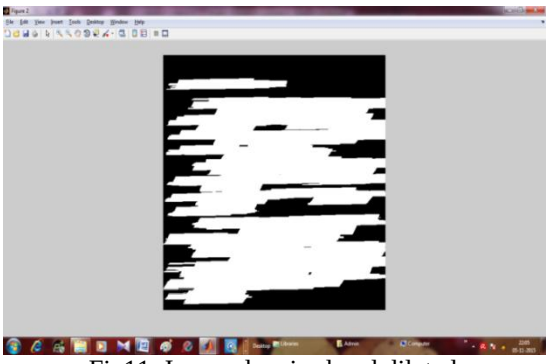

Fig11: Image denoised and dilated

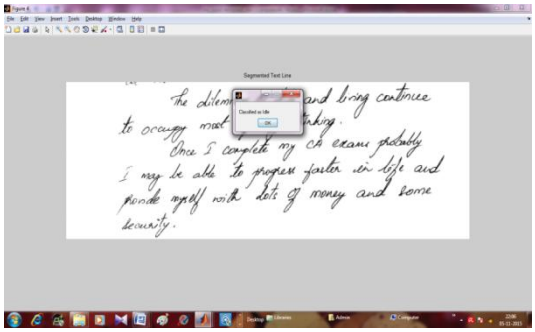

Figure 12:Mr.Mittals writing classified as ideal

VII. CONCLUSION

The current work is carried considered single lines and multiple lines for better assessing the personality trait. The training was done based on SVM. The network was trained for unknown samples of celebrities. The results were 94.4%. Samples were collected from different people at different point of time to make the network robust and results invariant.

IX. FUTURE WORKS

This work can be extended for considering other handwriting features such as margins, spaces between characters and so on to better judge the personality.

ACKNOWLEDGEMENTS

I would like to thank the Almighty above all then my parents my husband my kids and my guide for giving me encouragement in all kind of endeavours directly or indirectly. I further extend my thanks to the famous psychologist Gordon Allport.

REFERENCES

[1]. Jorge Azorín-López, Marcelo Saval-Calvo, Andrés Fuster-Guilló, José García-Rodríguez,: Human Behaviour Recognition Based On Trajectory Analysis Using Neural Networks, 2014 , In Proc IEEE Intl. Conf 2015.

[2]. Imran Ali Mirza,Shainila Mulla, Rishit Parekh, Satej Sawant, Krishna Mohan Singh : Generating Personalized Job Role Recommendations For The IT Sector Through Predictive Analytics And Personality Traits, 2015, In Proc. IEEE Int. Conf. On Technologies For Sustainable Development.

[3]. Charlesh.Trafford And Rand H.Nelson, Peterson : Directed Handwriting Handwriting A Complete Guide To Instruction Teaching Physical Patterns For Reading And Writing Fluency ISBN 1-890666-3 5-1, 2003.

[4]. Syeda Asra, Dr.Shubhangi D.C :Personality Trait Identification Using Unconstrained Cursive And Mood Invariant Handwritten Text ,2015, In Intl Journal Education And Management Engineering, Volume 5 Ppl 20-31

[5]. Amit Kr. Yadav, Abdul Azeem, Akhilesh Singh, Hasmat Malik : Application Research Based On Artificial Neural Network(ANN) To Predict No Load Loss For Transformer's Design, 2011,IEEE Intl Conf On Communication Systems And Network Technologies 2013.

[6]. Fereshteh Falah Chamasemani , Yashwant Prasad Singh : Multi-Class Support Vector Machine (SVM) Classifiers – An Application In Hypothyroid Detection And Classification , 2011. In IEEE Intl Conf On Bio-Inspired Computing: Theories And Applications, 2011.

[7]. José L. Vásquez, Carlos M. Travieso And Jesús B. Alonso: Using Calligraphies

Features For Off Line Writer Identification, IEEE Intl Conf. 2013.

[8]. Jun Zhang : Image Segmentation Based On 2D Otsu Method With Histogram Analysis ,IEEE Intl Conf. On Computer Science And Software Engineering 2008.

[9]. Wangsheng Yu, Xiaohua Tian ; Zhiqiang Hou ; Anqi Huang : Region Edge Histogram: A New Feature For Region-Based Tracking In Procd Signal Processing (ICSP), 2014 12th International Conference On Oct 2014.

[10]. Chapel GN,Daruwala RD,Gofane MS: Comparisions Of Robert, Prewitt, Sobel Operator Based Edge Detection Methods For Real Time Uses On FPGA.In Proc Intl Conf On ICTSD,Feb 2015.

[11]. Fereshteh Falah Chamasemani , Yashwant Prasad Singh Multi-Class Support Vector Machine (SVM) Classifiers – An Application In Hypothyroid Detection And Classification , 2011. In IEEE Intl Conf On Bio-Inspired Computing: Theories And Applications, 2011.

[12]. Fereshteh Falah Chamasemani, Yashwant Prasad Singh: Multi-Class Support Vector Machine (SVM) Classifiers – An Application In Hypothyroid Detection And Classification. In Intl Conf IEEE On BIC ,Oct 2011.