WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Personality Detection Using Signature Analysis

[1] D. Yashawini, [2] M. Harshitha, [3] Mounika, [4] B. Deepika Reddy, [5] Pranavi

[1,2,3,4,5] Dept. Computer Science & Engineering, Santhiram Engineering College, Nandyal, India

ABSTRACT

Personality is a unique thing that everyone has. Personalities show how a person acts, both in daily life and at work. Tracking a person's personality is important for the company, many companies conduct tests to get the personality of the prospective employee in accordance with company criteria. Graphology is a technique to analyze one's personality through handwriting. Deep learning is a development from machine learning. One of the most popular Deep Learning methods for image processing is Convolutional Neural Networks (CNN). Combining graphology approach and machine learning can create a computation system that can identify someone's personality automatically. This research is aimed to identify someone's personality using graphology approach and deep learning which is CNN. This research is using structural analysis with margin, dot structure, disconnected streak and separated signature features. While for symbol analysis using CNN for curved start, end streak, upper streak, middle streak, shell signature and underline features. The results show that the accuracy of structural analysis approach was up to 78.20-96.20% and the accuracy of symbol analysis with CNN was up to 98.46%. Choosing better configuration and optimization method can improve high accuracy.

Keywords: Personality, Handwritten Signature, Graphology, Structural Analysis and Convolutional Neural Networks.

1. Introduction

Personality is a unique thing that everyone has. Personalities show how a person acts, both in daily life and at work. Apart from the personality that is owned, we can "categorize" someone as a person who is conscientious or reckless, swift or sluggish in carrying out activities, both professional and daily activities. Tracking a person's personality is important for the company, many companies conduct tests to get the personality of the prospective employee in accordance with company criteria. Besides being used for employee recruitment, personality information is also useful in academics, mental counseling, and forensic information. One way to analyze someone's personality through handwriting is called Graphology. Graphology is a technique to analyze one's personality through handwriting, handwriting analysis projects a description of behavior in the areas of social skills, achievements, thinking styles, and work habits (Rahiman M *et al.*, 2013). Handwriting is a neuromuscular movement that is typically associated with brain patterns that occur unconsciously, scratches or handwriting are often referred to as "brain writing". The resulting hand strokes define and distinguish one stroke from another, which is associated with personality traits (Sen and Shah, 2018), (Djamal *et al.*, 2013). Both structural and symbol analysis can be used to analyze personalities based on the signature.

Current technology allows a machine to recognize image patterns, signals automatically, and therefore graphology techniques can be used in personality analysis considering the analysis is done on the signature image. Previous research has been carried out on personality analysis through signature (Djamal et al., 2013), (Lokhande and Gawali, 2017). Research (Djamal et al., 2013) carried out personality identification from signature using nine features, four features including margin, dot structure, separated signature and disconnected streaks are analyzed using multi-structure algorithm which result an accuracy of 87-100%. While for the other five features are analyzed using Artificial Neural Networks (ANN) including curved start, end streak, shell, middle streak and underline get an accuracy of 56-75%. The identification are built using Multilayer Perceptron (MLP) architecture, it obtained five networks for each features. It consist of 828 neurons in input layer, 10 neurons in hidden layer, and 2 neurons in output layer except for curved start features. While the network for shell feature consist of 2848 neurons in input layer, 10 neurons in hidden layer, and 2 neurons in output layer, this configuration resulted low accuracy because the number of neuron in hidden layer is less than neuron in input layer. The use of ANN on large input dimensions causes a long processing time, consideration of using a different method and configuration can improve more accuracy and faster time processing. Other studies analyze signatures to identify authentic or forged signatures include using ANN (Rehman et al., 2018), (Chandra and Maheskar, 2016), Histogram Oriented Gradient (HOG) (Hasan Abbas et al., 2018), and Backpropagation Neural Networks (BPNN) (Inan and Sekeroglu, 2018).

Deep learning is a development from machine learning which offers better performance and high accuracy using extensive and interconnected feature extraction, however it needs a decent hardware and more training data. One of the most popular Deep Learning methods for image processing is Convolutional Neural Networks (CNN). Some research using CNN for personality identification from handwriting (Fatimah et al., 2019), (Valdez-Rodr´ıguez et al., 2019), personality identification from palmprint (Ariyanto et al., 2019). Other studies using CNN for signature verification (Mohapatra et al., 2019), (Sronothara and Hanmandlu, 2018), (Wencheng et al., 2018), (Pinzón-Arenas et al., 2019).

This study proposed to identifying person's personalities from the signature image by structural and symbol analysis approach. The structural analysis features are based on (Djamal et al., 2013), while for symbol analysis classified using CNN methods. In this study we add features that not used previously which is four rules in margin features, upper stroke and many underlining that analyze with CNN method.

2. Proposed Method Graphology

Structural analysis for graphology feature are based on (Djamal et al., 2013), include margin, dot structure, separated signature, disconnected streak and four addition rules on margin feature can be seen on Table 1

Table 1. Structural Analysis Features Description

No	Description	Type	Example
1	Margin: Negligence; Pessimist; Cheerful person; Depression, feel strange	8	 
2	Dot Structure:Stable establishment, Distrustful	2	
3	Separated Signature: bitterexperience in the past	2	
4	Disconnected Stroke: Discouraged, afraid to make decision	2	

While for symbol analysis using CNN method include curved start, end streak, shell signature, middle streak, underline and addition feature which is upper streak, and many underline in underline feature. The symbol analysis description can be seen on Table 2.

Table 2. Symbol Analysis Feature Description

No	Class	Description	Type	Example
1	Curved Start: - Smooth curved - Backward curved - Sharp curved	Good experience in the past. careful mind. friendly.	3	
2	Ending Stroke - Ascending ending stroke - Descending ending stroke	Confidence, open minded. Less enthusiastic, realistic. Feel hopeless.	3	
3	Shell Signature: - Encircled signature.	Introvert	1	

4	Middle Stroke - Middle stroke signature.	Possessive.	1	
5	Underline - Underline signature - many underline signature.	Good leadership, thinking uniquely Doubt his own public worth.	2	
6	Upper Stroke - Upper stroke signature.	protectiveness, leadership traits.	1	
7	Others	Apart from eleven class type above, no personalitytraits defined.		-

Data Acquisition

Signature data for CNN Analysis are captured using smartphone camera, we asked random people to write their signature 10 times each. Afterward we categorized the signatures according to graphology features, each features has 1000 datasets of signature image so that 11000 data are collected for training. The data stored in JPG format and already threshold with size of 240x240 pixels. Afterward, for testing the personality identification, we collected signature images from 50 authors who were asked to write their signature in accordance with the format provided, which is to write down the signature in a formatted form on A4 sized paper. There is no specific analysis that considers the suitability of the features in the existing graphology table with the built model. The computational models built for personality identification based on signature images consist of preprocessing, segmentation and feature extraction in order to extract features from existing images, which are shown in Figure 1. Therefore, these features can show a person's personality from various points of view.

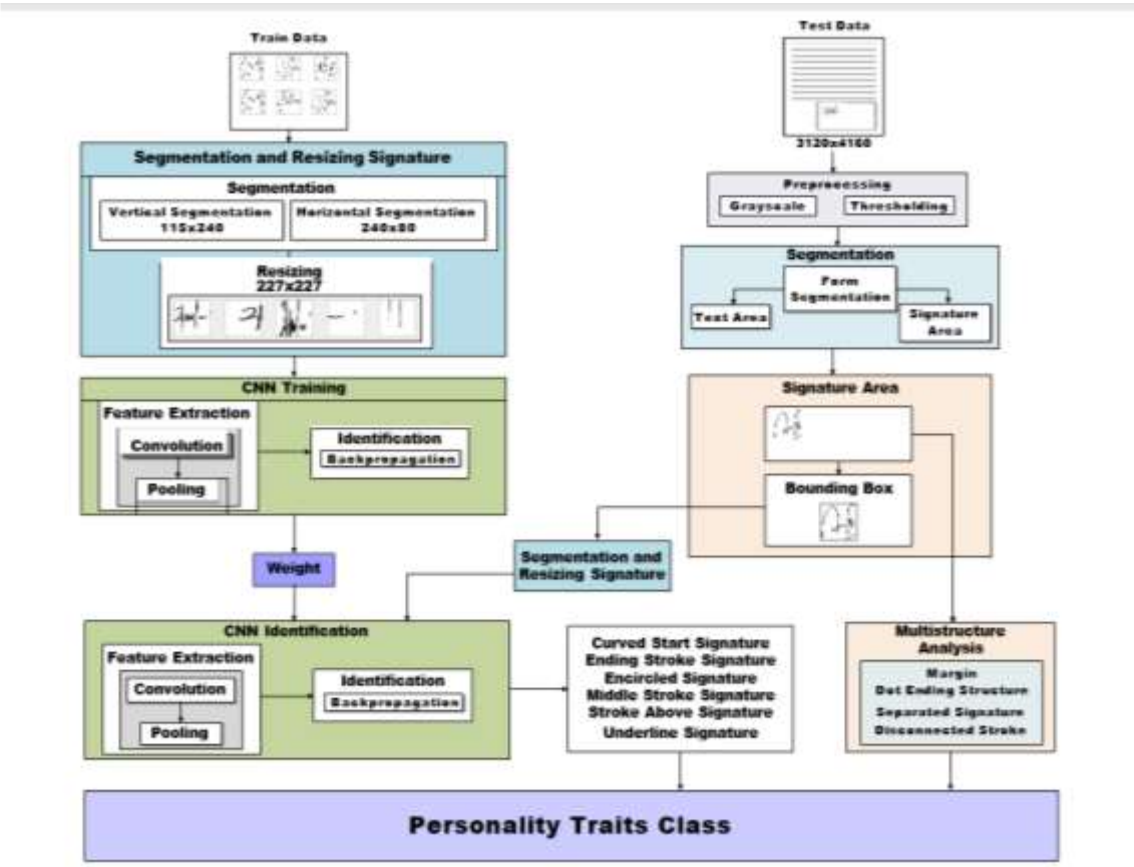


Figure 1. Personality Identification Model

Pre-processing and Segmentation

a. Pre-processing

Preprocessing is done to improve the quality of the image and bring up features that can be extracted, pre- processing is done in two stages, including grayscale and threshold. Grayscale is the stage of changing the color scale of an image from RGB to gray (1), then the color of the image is changed to black or white at the threshold stage (2). Images that have gone through the pre-processing stage will be easier to process during segmentation and structural analysis due to significant differences in pixel values. Where (x, y) represent the pixel coordinates.

$$grayscale = 0{,}299R + 0{,}587G + 0{,}114B \quad (1)$$

$$f\,(x, y) = \{0, \qquad f1(x, y) < 128$$

$$0 \qquad 1, \qquad f1(x, y) \geqslant 128 \quad (2)$$

b. Segmentation

Segmentation aims to eliminate parts of the image that are not needed to speed up the computing process. Segmentation begin with form segmentation, this phase is done to eliminate text area on the image so we get the signature part that is in a box. A signature box is created so that the position on the signature can be analyzed with structural analysis for margin features. The segmentation method are based on (Djamal et al., 2013).

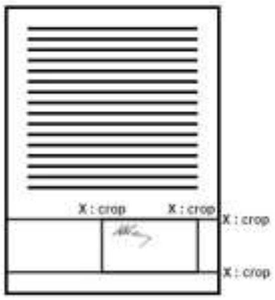


Figure 2. Form Segmentation

After the segmentation form is done, then the signature in the box is given a bounding box so that the signature image is obtained. Analysis of the structure of the dot structure, separated signature, disconnected stroke features then can be carried out, whereas for CNN analysis the signature image of the bounding box results is resized to 240x240 pixels, then vertical and horizontal segmentation are performed.

Vertical segmentation is done to identify the initial features of curves and final streaks, and horizontal segmentation is done to identify the features of upper stroke, middle stroke and underline. As for the shell signature feature, the bounding box image signature is only resized to a size of 227x227 pixels, similar to the results of vertical segmentation and horizontal segmentation are resize to 227x227 pixels.

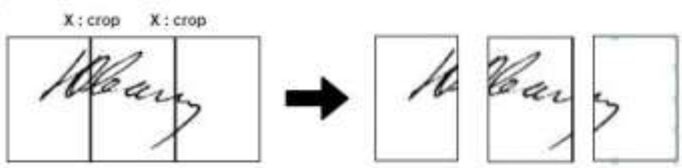


Figure 3. Vertical Segmentation

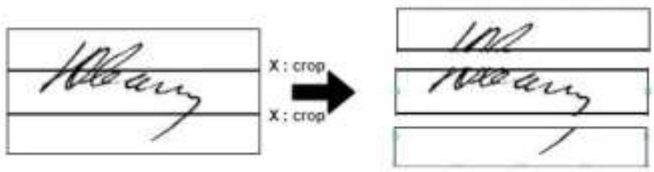


Figure 4. Horizontal Segmentation

Structure Analysis

Structural analysis is done to identify features without CNN training. Structural analysis can be done after the form segmentation proses.

a. Margin

To analyze the margin feature, a box is provided where the signature is written, this box helps to locate the signature. Signature position search is performed by comparing positions on X which is initialized for left or right, and Y for up or down. Margin position includes left, upper left, bottom left, right, upper right, bottom right, upper inclines, bottom inclines.

b. Dot Structure

Analyzing dot structure in signature is done by tracing black pixel that separated with withe pixels. Dot structure in signature usually located at the ending of signature with dimension about 5-20 length, and 5- 15 width.

c. Separated Signature

To analyze separated signature, the first thing to do is to find the coordinate of main signature. After getting the coordinates of the main signature, then search for a separate coordinate point starting from the end point of the main signature coordinates.

d. Disconnected Streak

Analyzing disconnected in signature is done by tracing separated black pixels in vertically and horizontally. Stroke detection is done as same as dot structure, however the strokes are sized about 20-40 pixels long, and 5-15 pixels wide.

Convolutional Neural Networks

CNN is one method of deep learning that is usually used in image processing, in this study we adapted AlexNet architecture, AlexNet is CNN architecture designed by Alex Krizhevsky (Krizhevsky et al., 2012). This research adapt the AlexNet architecture with 5 convolution layer and 1 fully-connected layer. While the classification process for the fully-connected layer uses Multilayer Perceptron (MLP) architecture. The CNN architecture can be seen on Figure 5.

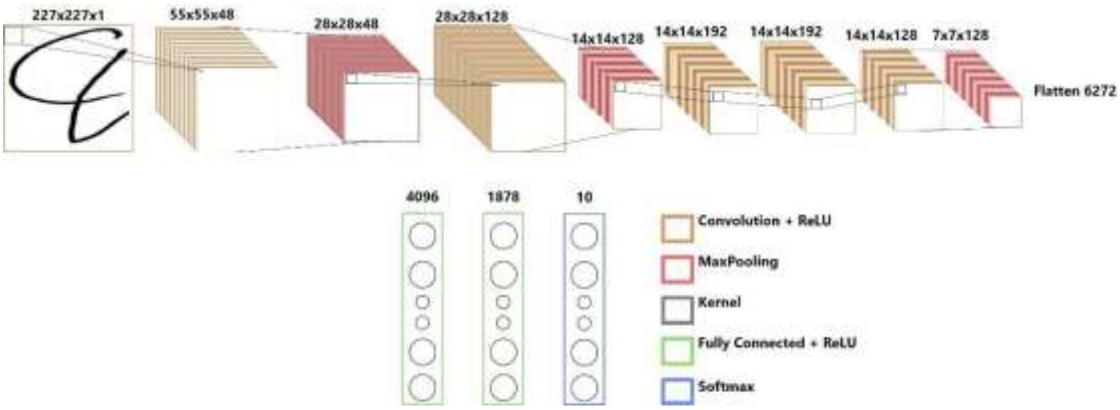


Figure 5. Convolutional Neural Networks Architecture

a. Feature Extraction

On this stage, there is a transformation process from the input to find several vector shaped features, before entering the identification layer. This feature extraction stage consist of convolution layer and pooling layer.

1. Convolution

The convolution process converts the image into a feature map by performing a dot value product input operation with a kernel filter (Ariyanto et al., 2019). Convolution has some hyper-parameters, including stride, depth and padding value denoted in (3).

$$FM(l,ml)_{(il,jl)} = f\left(\sum_{rl=0}^{kh-1} \sum_{cl=0}^{kw-1} C(l,ml)_{(rl,cl)} * FM(l-1)_{((rl+il-1),(cl+jl-1))}\right) \quad (3)$$

Where l is index of input/layer, ml is index of map on l layer, il is row index feature map l layer and jl is row index feature map l layer. kh is height of kernel filter, kw is length of kernel filter, rl is length of kernel layer l, cl is height of kernel layer l. C is the kernel.

After convolution process is finished and FM is obtained, the next step is activation process using Rectified Linear Unit (ReLU) to normalize negative value to zero, so that all the value in FM is a positive value. The ReLU function is denoted in (4).

$$f(x) = \max(0, x) \qquad (4)$$

2. Pooling

Pooling or downsampling is useful to reducing the dimension of FM to minimize computational loads. Pooling will take value of feature map in the area of specified size and stride. This study using Max Pooling which take highest value in feature map, as shown in Figure 6.

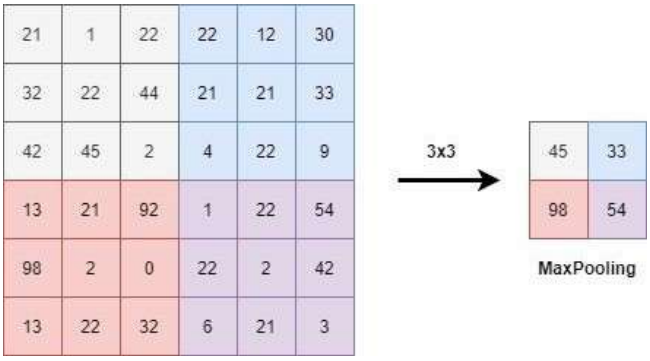


Figure 6. MaxPooling

b. Classification Layer

Classification layer has several configurations, and the most commonly used is MLP (Wijaya et al., 2018). Feature map from feature extraction stage is matrix, then converted into a flatten vector as an input of the layer.

1. Softmax Activation

The Softmax function is used to calculate the probability of each target class for all possible target classes and will help to determine the target class for a given input (Dyrmann et al., 2016). Softmax function is denoted in (6) where w is the weight value and and m is the number of units in the output layer.

$$(y_k) = \frac{e^{x^T w_k}}{\sum_{i=1}^{m} e^{x^T w_i}} \qquad (6)$$

2. Cross-Entropy Loss Function

Loss function is used to measure how much deviation occurs between the classification results and the actual data. This study use Cross-Entropy Loss Function (Dyrmann et al., 2016) denoted in (7). Where m is number of units in output layer, p is label class and $\hat{p}_i$ is result after Softmax activation.

$$Cross\ Entropy = -\sum_{i=1}^{m} p_i \log \hat{p}_i \qquad (7)$$

c. Optimization Method

Optimization method is used to correct weight. This research use Stochastic Gradient Descent (SGD) and Adaptive Learning Rate (Adadelta). Both method are intended to revise the weight of generalization, to minimize the error value or deviation between actual output and label. The difference between SGD and Adadelta is that SGD trains the network using a random network sample to correct the weights. This weighting improvement is done in stages which can cause stalled repairs to the minimum local iteration (Yang and Yang, 2018). Adadelta is a development of SGD which has techniques for adapting weight improvement so that it runs dynamically.

3. Result and Discussion

Result and discussion are divided into a structural analysis and symbol analysis. Structural analysis features include margin, dot structure, separated signature, disconnected streaks. The symbol analysis feature include curved start, end streak, uppor stroke, middle stroke, underline.

Structural Analysis Result

Testing with structural analysis is carried out using 50 signature samples and obtained the result in Table 3. The result show that dot structure gained 78.20% accuracy, this occurred because the analysis rules for dot structure is not match with the dot dimension on actual data, same happened with separated signature. The image quality, pre-processing and the segmentation process affects the result of each feature.

Table 3. Structural Analysis Result

Feature Category	**Accuracy(%)**
Margin	96.20%
Dot Structure	78.20%
Disconnected Streak	78.30%
Separated Signature	80.20%

Symbol Analysis Result

Symbol analysis is done using CNN method symbol analysis result contains of the training model for symbol analysis features. There are two test scenario which is learning variety comparison, and optimization method comparison. The data is split into 8000 train data and 2000 test data. The model trained with GPU, the device specifications used is Core i5 9300H, 8GB RAM, NVIDIA GeForce GTX 1050TI. There are three type of CNN configuration that used in this research, the configuration is shown in Table 4.

Table 4. Convolutional Neural Networks Configuration

Layer	**Configuration 1**	**Configuration 2**	**Configuration 3**
Input	227x227x1	224x224x1	224x224x1
Conv-1	55x55x48	54x54x48	54x54x96
Max Pool-1	28x28x48	26x26x48	26x26x96
Conv-2	28x28x128	262x26x128	22x22x256
Max Pool-2	14x14x128	12x12x128	10x10x256
Conv-3	14x14x192	12x12x192	8x8x384
Conv-4	14x14x192	12x12x192	6x6x384
Conv-5	14x14x128	12x12x128	4x4x256
Max Pool-3	7x7x128	5x5x128	1x1x256
Flatten	6272	3200	256

Layer	**Configuration 1**	**Configuration 2**	**Configuration 3**
Fully Connected-1	4096	1595	246
Fully Connected-2	1878	1595	-
Output Layer	10	10	10

a. Learning Rate Variety Comparison Result

The first test scenario is done by comparing learning rate variety both on SGD and AdaDelta with 3 CNN configuration. The learning rate variety consist of 0.001, 0.002, 0.010, 0.020, 0.100. The best result is gained from SGD with 0.002 learning rate and configuration 2 of CNN, it is shown that more complex network can carried out the feature that can improve accuracy. While for AdaDelta best accuracy shown with configuration 2 of CNN and 0.002 learning rate. This result show that the number of kernel effects on producing the number of input in classification.

Table 5. Result of Learning Rate Variation on SGD

Learning Rate	Configuration 1	Configuration 2	Configuration 3
0.001	98.40%	97.55%	83.95%
0.002	98.46%	98.35%	84.20%
0.010	10.25%	10.50%	7.35%
0.020	9.90%	10.45%	8.34%
0.100	9.35%	8.01%	6.12%

Table 6. Result of Learning Rate Variation on AdaDelta

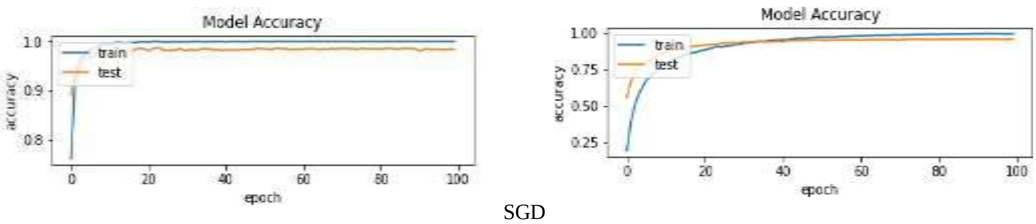
Learning Rate	Configuration 1	Configuration 2	Configuration 3
0.001	94.40%	90.55%	80.95%
0.002	95.80%	91.35%	83.20%
0.010	10.25%	8.50%	7.35%
0.020	9.90%	7.45%	8.34%
0.100	9.35%	7.01%	6.12%

b. Optimization Method Comparison Result

Testing the effect of optimization models is done to find out which optimization model is more suitable for personality identification based on the signature image. This optimization model is used to update the weights when training. The optimization models tested are Stochastic Gradient Descent (SGD) and Adadelta. Accuracy results from testing both optimization models can be seen in Table 7, Figure 7 and Figure 8.

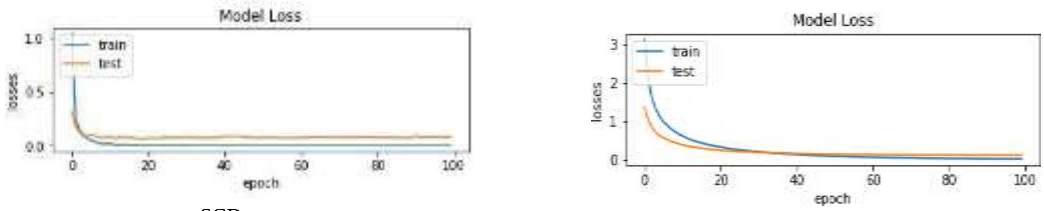
Table 7. Optimization Method Comparasion Result

Experiment	Data Train		Data Test	
	Accuracy	Loss	Accuracy	Loss
SGD	100%	0.0024%	98.46%	0.079%
Adadelta	100%	0.0013%	95.80%	0.121%



SGD Adadelta

Figure 7. Model Accuracy



SGD

Adadelta

Figure 8. Model Loss

Figure 8. shown AdaDelta loss is smaller than SGD, the loss of AdaDelta is 0.013% of data train while SGD loss is 0.024% for data train, however the accuracy gained with SGD is higher than AdaDelta, shown in Figure 7. AdaDelta show stable learning or validation, whereas SGD shows more volatility in the training data. In this study the Adadelta optimization model tends to be more stable and constant, also does not experience a significant increase in loss compared to SGD which tends to be unstable, because when conducting the SGD training process only uses one or several parts of the training data chosen randomly, causing a minimum occurrence local. Adadelta also has a unique change in new learning parameters by calculating the average of the error values. This is what helps Adadelta provide the most optimal value to accelerate convergence as in Figure 8. Creating better configuration can improve high accuracy of learning, the overall accuracy of identification is better than previous research which obtained 56-75% accuracy, although there are differences in features (Djamal et al., 2013).

4. Conclusion

Personality identification from handwritten signature can provide several reviews, based on the features used. From 50 sample margin feature, it gained 37 samples correctly identified. The error is caused by miss calculating on eight rules of margin features. For dot structure and disconnected streak only gained 78% accuracy, it is caused by miss calculating dimension for dot structure and the disconnected streak. While for separated signature it gained 34 samples correctly identified and gained 80.20% accuracy.

The structural analysis produce 78.20-96.20% accuracy while for symbol analysis using CNN method produce 98.46% accuracy. CNN configuration and optimization model determine accuracy in symbol features to define personality. SGD optimization provide high accuracy which is 98.46% compared to Adadelta which is 95.80%. In this study the Adadelta optimization model tends to be more stable and constant, also does not experience a significant increase in loss compared to SGD which tends to be unstable, because when conducting the SGD training process only uses one or several parts of the training data chosen randomly, causing a minimum occurrence local. Adadelta also has a unique change in new learning parameters by calculating the average of the error values. This is what helps AdaDelta provide the most optimal value to accelerate convergence.

References

Ariyanto, A., Djamal, E. C. and Ilyas, R. (2019) 'Personality Identification of Palmprint Using Convolutional Neural Networks', Proceeding - 2018 International Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity, SAIN 2018. IEEE, pp. 90–95.

Chandra, S. and Maheskar, S. (2016) 'Offline signature verification based on geometric feature extraction using artificial neural network', 2016 3rd International Conference on Recent Advances in Information Technology, RAIT 2016, pp. 410–414.

Djamal, E. C., Darmawati, R. and Ramdlan, S. N. (2013) 'Application image processing to predict personality based on structure of handwriting and signature', Proceeding - 2013 International Conference on Computer, Control, Informatics and Its Applications: 'Recent Challenges in Computer, Control and Informatics', IC3INA 2013, pp. 163–168.

Dyrmann, M., Karstoft, H. and Midtiby, H. S. (2016) 'Plant species classification using deep convolutional neural network', Biosystems Engineering. Elsevier Ltd, 151(2005), pp. 72–80.

Fatimah, S. H. et al. (2019) 'Personality features identification from handwriting using convolutional neural networks', 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2019, pp. 119–124.

Hasan Abbas, N. et al. (2018) 'Offline Handwritten Signature Recognition Using Histogram Orientation Gradient and Support Vector Machine', Journal of Theoretical and Applied Information Technology, 96(May), pp. 2075–2084.

Inan, Y. and Sekeroglu, B. (2018) 'Signature Recognition Using Backpropagation Neural Network', 13th International Conference on Theory and Applications of Fuzzy Systems and Soft Computing, 896, pp. 256–261.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) 'ImageNet Classification with Deep Convolutional Neural Networks', Neural Information Processing System (NIPS 2012), pp. 1–9.

Lokhande, V. R. and Gawali, B. W. (2017) 'Analysis of Signature for the Prediction of Personality Traits', Proceedings - 1st International Conference on Intelligent Systems and Information Management, ICISIM 2017, 2017-Janua, pp. 44–49.

Mohapatra, R. K., Kumar, S. and Kedia, S. (2019) 'Offline Handwritten Signature Verification using CNN inspired by Inception V1 Architecture', International Conference on Image Information Processing (ICIIP 2019), pp. 1– 5.

Pinzón-Arenas, J. O., Jiménez-Moreno, R. and Pachón-Suescún, C. G. (2019) 'Offline signature verification using DAG-CNN', International Journal of Electrical and Computer Engineering, 9(4), pp. 3314–3322.

Rahiman M, A., Varghese, D. and Kumar G, M. (2013) 'HABIT: Handwritten Analysis Based Individualistic Traits Prediction', International Journal of Image Processing (IJIP), (7), p. 209.

Rehman, A. U. et al. (2018) 'Offline Signature Recognition and Verification System Using Artificial Neural Network',

University of Sindh Journal of Information and Communication Technology, 2(1), pp. 73–80.

Sen, A. and Shah, H. (2018) 'Automated handwriting analysis system using principles of graphology and image processing', Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017, 2018-Janua, pp. 1–6.

Sronothara, A. B. and Hanmandlu, M. (2018) 'Off-line signature verification using CNN', International Journal on Future Revolution in Computer Science & Communication Engineering, 4(9), pp. 85–90.

Valdez-Rodr´ıguez, J. E., Calvo, H. and Felipe-River´on, E. M. (2019) 'Handwritten Texts for Personality Identification Using Convolutional Neural Networks', International Conference on Pattern Recognition (ICPR) 2018, 11188, pp. 66–75.

Wencheng, C. et al. (2018) 'Offline Chinese signature verification based on AlexNet', Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST, 219, pp. 33–37.