Songklanakarin J. Sci. Technol.
41 (1), 136-143, Jan. – Feb. 2019

http://www.sjst.psu.ac.th

Original Article

Applications of time-series analysis: A case study on the development of a predictive model for handwritten signature

Alpana Deka[1] and Lipi B. Mahanta[2*]

[1] *Department of Statistics, Gauhati University, Guwahati, 781014 India*

[2] *Institute of Advanced Study in Science and Technology, Guwahati, 781014 India*

Received: 19 July 2017; Revised: 22 September 2017; Accepted: 5 October 2017

Abstract

Handwritten signature represents a person's identity. Although overall patterns among the signatures of same person remain same, there can appear natural variations because two or more signatures of same person written within a moment and keeping a sufficient time gap, cannot be exactly same. These natural variations result in intrapersonal variations. In the present study, signature samples were collected from each participant under different situations of body position, paper texture, paper position etc. to successfully capture the intrapersonal variations. Two features, namely area and height-width ratio (HWR) were extracted for each signature using appropriate image processing techniques. These features were then modelled to the Single Exponential Smoothing Time Series Technique as well as our developed methodology to predict the variations. Using this technique the Positive Predictive Values (PPV) and False Rejection Rate (FRR) for both these features were found to be 88%, 12% and 95.78%, 4.22% respectively.

Keywords: offline signature analysis, image processing, time series, single exponential smoothing technique, predictive modeling

1. Introduction

The term "Forecasting" means prediction of future data by studying the pattern of present and past data or historical data. Depending on the length of the forecast horizon to which the prediction should be made, the forecasting may be of three types: short-term, medium-term and long-term (Fotak, Baca, & Koruga, 2011). By studying the historical data, the concept of forecasting can be applied to handwritten signatures also. Due to several factors or conditions, intrapersonal variations appear among the signatures. The most common such variations are: written quality of letters, connectivity or continuity and spacing between letters or words, pen pressure, skew and slant angle, misplacement of cross and dot marks, stroke marks etc. (Azzopardi, 2006; Bertino & Bertino, 2011). The factors responsible for the variations are body position of the person, writing material, purpose of signing, surrounding environment, old age and other ailment (eye) conditions, physical and psychological state of the person etc. (Gonzalez, Woods, & Eddins, 2010; Houck, & Siegel, 2015). We have considered the above -mentioned variations as far as possible to deal with the intrapersonal variations in our proposed system of forecasting. Due to the presence of such variations, a handwritten signature, although genuine, gets rejected often in an online checking format. All these causes are completely genuine and the signatories involved often cannot make a remedy in the signature. This causes harassment to both the sides involved in the process involving the signature. This situation motivated the present study. Here, it is noted that a lot of research work on handwritten signature has been done for proposes of verification of signatures till now. We know that the concept of forecasting is widely applicable in different areas namely business related predictions, governmental budgets, policies

*Corresponding author
Email address: lipimahanta@yahoo.co.in;
lbmahanta@iasst.gov.in

and planning, and other areas like weather as well as temperature etc. However, signature forecasting may be a new application area which has not been given great importance till now. In this paper we try to incorporate this new idea. Current research trend in the field of signature verification systems is focused mainly on identification of forgeries, whether simple, random or skilled. (Bertolini, Oliveria, Justino, & Sabourin, 2010; Scheidat, Vielhauer, & Dittman, 2009) Also different studies deal either with online or offline signatures. These methods apply different techniques of image processing and pattern recognition techniques to achieve the goal. A review of these studies were presented by Mahanta and Deka (2013). Our proposed system focuses on a new application area of forecasting in handwritten signature, which is the main contribution of the paper.

2. Methodology

In our proposed system, we have adopted the following methodology as given in Figure 1, to forecast signatures of a person.

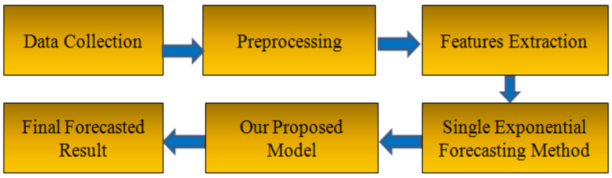


Figure 1. Proposed methodology of signature forecasting system.

2.1 Data collection

Here, genuine signature samples were collected from each of 15 participants over a period of one year and 3 months, maintaining a 3 - month time gap between each stage of collection. Thus the task of sample collection was completed in 6 stages. During the 1st stage, each participant was asked to sign 7 signatures at a time on an "A4 size print paper (manufactured by JK Paper Ltd)" in sitting position, since during a signing process, sitting is a person's most usual position. These 7 signatures were applied as a training set for each of the respective participants. Then during each of the remaining 5 stages, the participants were asked to provide 2 samples at a time under different situations. Most common situations are explained below:

a) Body Position: Here, a participant was asked to sign by maintaining three different body positions.

- Sitting properly: This is the usual signing condition (sitting in a chair) without any uneven body position. Here, the person completes the signing process by keeping an exercise book in between the signing paper and a wooden table such that any roughness of the table cannot affect the signature.
- Half standing: Here, a person is allowed to sign on a paper in half-standing position such that no portion of his body touches any sitting materials.
- On wall: Sometimes, it happens that there is no table or no spaces available a table for a person to sign. During that situation, usually the person prefers to sign by holding the signing paper on a wall. In the proposed system, a person is allowed to keep an exercise book between the signing paper and wall such that any uneven structural form of the wall cannot affect the signature.

b) Writing surface: Often, the writing surface upon which signing paper is kept do not smooth. To deal with such a situation, each sitting participant was provided the same type "rough textured hard board" to keep between the signing paper and the table.

c) Surrounding area not free: A situation may also arise, where a table is full of books with other materials and only little space remain available, such that a sitting person can sign on a paper by keeping an exercise book between the paper and table. Here also the same condition was created to each of the participant to perform the signing process.

d) Normal Paper: During data collection, the above stated situations were completed with A4 size print paper. Owing to different paper qualities, the natural variations arise, therefore signatures are also collected on normal papers (of rougher texture and which are usually used to copy any notes) from each sitting person.

e) Other factors: Besides the above mentioned situations, other factors which were taken into consideration during the time period of samples collection are noted below:

First of all, during the signing processes, no restriction was imposed on the type of pen with which he/she would sign. Secondly, as the sample collection process was spread over one year, variations in atmospheric and environmental factors (like noisy or quiet surroundings, temperature, humidity etc.) and health and psychological states, which may affect the signing process were covered.

Thus in our proposed system, the sizes of training and testing sets, consisting of only genuine signatures, are 105 and 900 respectively.

2.2 Image preprocessing

Next, we have applied a few image preprocessing techniques (namely complement binarization, removal of redundant bordering components, adjustment of signature's spaces etc.) on each of the signatures to improve the quality of digitized signature image such that a noise - free signature would be isolated (Jain, & Malehorn, 2005; Karouni, Daya, & Bahlak, 2011).

2.3 Features extraction

"Features" is a set of inherent characteristics that represent an object. In handwritten signatures, such inherent characteristics are analyzed to study their behavioral nature. The features that are most used for verification of forgery are Area and Height-Width Ratio (HWR), Height-Width Ratio (HWR), Normalized area (NA), Maximum vertical projection (MVP), Maximum horizontal projection (MHP) and Sum of four local Normalized Areas (SLNA). In our proposed system, we analyzed two features, namely Area and Height-Width Ratio (HWR), to be further applied to forecast the future behavior of signatures (Kisku, Gupta, & Sing, 2010; Mahanta,

& Deka, 2013; Makridakis, Wheelwright, & Hyndman, 2008). These two features are most commonly used and also add maximum to the analysis of a signature.

2.4 Image area (IA)

Image area was obtained by summing the "on pixels" present in the signature. In our case as displayed in Figure 4, the term "on pixels" represents the white foreground pixels of the image. If A_{ij} (where i = 1,2,… … ,m and j = 1,2,… … ,n) indicates elements that are present on foreground of the signature, then to find the area, we will first searched those pixels whose values are 1 i.e. $A_{ij} == 1$ and then summed them. Mathematically, the area can be calculated as given below:

$$\sum_{i=1}^{M} \sum_{j=1}^{N} (A_{ij})$$

2.5 Height to width ratio (HWR)

We calculated height and width of the signature by following number of rows and columns (or dimension) confined in the image as given below:

$$\text{Height} = h_{last} - h_{first} + 1$$
$$\text{Width} = v_{last} - v_{first} + 1$$

Here, h_{first} , h_{last} are topmost, bottommost rows and v_{first} ,v_{last} are leftmost, rightmost rows of the signature, respectively.

Finally, we evaluated the desired ratio by taking the ratio between the computed values of height and width of the signature.

In future, we will extend our work to other features like maximum horizontal projection, maximum vertical projection etc.

2.6 Single exponential smoothing technique

Usually, in almost all applications of time series forecasting model (like weather, population and crop production etc.), it is observed that historical data are already available and can be collected from their respective resources. As a result, we can collect all the recorded historical data depending on desired time periods and hence it is not a time consuming process. But in the case of signatures forecasting, we do not have ready-to-use recorded historical data. Here, by maintaining sufficient time gap between every stage, signature samples have to be collected under different situations from respective signers. These extracted features can be used as historical data. Thus for signatures forecasting, the historical data collection step becomes more and more time consuming. In our proposed system, we took 7 samples as a training set for each person. The plotting of each such training set (Figure 2) shows that the data are horizontal or stationary, which means that there are no trends or seasonal components present in the data. Comparing the requirement of data for each of them with our availability of training data, the single exponential smoothing technique is found to be more appropriate as it does not require so much training data as in ARIMA and moving average models.

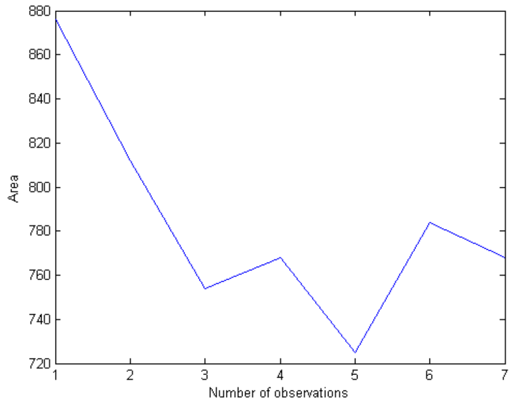


Figure 2. Plotted data of training set (feature: Area)

The single exponential smoothing technique is stated below (Shah, 2009; Altman, & Bland, 1994):

The basic notations are given by:

$F_{(t+1)}$: the next period for which the corresponding feature of a given signature to be forecasted.

α : the smoothing constant or smoothing weight that takes a value within the range of 0 and 1.

Y_t : the actual value of the corresponding feature of a given signature during the current period.

F_t : the forecast value of corresponding feature of a given signature during current period.

Then

$$\begin{aligned} F_{(t+1)} &= F_t + \alpha (Y_t - F_t) \\ &= F_t + \alpha Y_t - \alpha F_t \\ &= \alpha Y_t + (1-\alpha) F_t \end{aligned}$$

We have applied the single exponential smoothing technique with $\alpha = 0.1, 0.2, \ldots\ldots, 0.9$ and then, to measure the historical error, we have applied MAD (Mean Absolute Deviation) as given below:

$$MAD = \frac{\sum |Actual\ value - Forecast\ value|}{Total\ number\ of\ years\ for\ measuring\ forecast\ errors}.$$
$$= \frac{\sum |Y_t - F_t|}{n}$$

2.7 Proposed model

To find out middle or long-term forecasting, we have applied our proposed methodology on the result of short-term forecasting with a trial and error method such that the resulting outcome appears with an optimum range. The proposed model is given below:

Suppose,

x_1 denotes the resultant short-term signatures forecasting $F_{(t+1)}$.

x_L and x_U denote the lower and upper limits. Then

$$x_L = x_1 - \frac{x_1}{4}$$

$$x_U = 2 \times x_L$$

We can express the above equations as stated below:

$$Lower\ Limit = short\ term\ forecasting - \frac{short\ term\ forecasting}{4}$$

$$Upper\ Limit = 2 \times Lower\ Limit$$

Thus the resultant limit is

$$[x_L, x_U] = [x_1 - \frac{x_1}{4}, 2 \times x_L]$$

Or,

$$[Lower\ Limit, Uper\ Limit] = \begin{bmatrix} \left(short\ term\ forecasting - \frac{short\ term\ forecasting}{4}\right), \\ (2 \times Lower\ Limit) \end{bmatrix}$$

3. Results

After completion of sample collection, the samples were converted to digitized format with the help of a scanner. Some of the signatures of a participant collected in different situations are displayed in Figure 3 below:

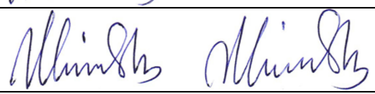
Situations	
Sitting properly	
Half Standing Position	
Paper kept “on wall”	

Figure 3. Signatures in different signing positions

The resultant images after preprocessing operations on signatures in ‘sitting properly’ position, as shown in Figure 3, are shown in Figure 4.

Figure 4. Application of pre-processing operations

Figure 5 denotes the height and width of the signature which is used to calculate the height-width ratio of a signature.

Figure 5. Height and width of a signature

Now, computed features of Area and HWR of the training and test set are given in Table 1 to 3.

Table 1. Extracted features from the training signature samples (1st stage of data collection).

Collected signatures	Area	Height-Width Ratio (HWR)
1	876	0.4410
2	812	0.4167
3	754	0.4190
4	768	0.4011
5	725	0.4231
6	784	0.4227
7	768	0.4475

Table 2. Extracted features of Area during stage 2-6 of data collection.

Situation	Sitting Properly	Half Standing	On Wall	Hard Copy	Surrounding Not Free	Normal Paper
Stage 2	840.5	811	892.5	811.5	822	877.5
Stage 3	844	797	843.5	837	754.5	975.5
Stage 4	891.5	792.5	880.5	795	936	809
Stage 5	821	797.5	771.5	794	722.5	776
Stage 6	695	640.5	755	665	777.5	728.5

Table 3. Extracted features of Height-Width Ratio during stage 2-6 of data collection.

Situation	Sitting Properly	Half Standing	On Wall	Hard Copy	Surrounding Not Free	Normal Paper
Stage 2	0.458	0.5354	0.5719	0.5175	0.4902	0.4448
Stage 3	0.4908	0.4587	0.5219	0.4756	0.4869	0.4092
Stage 4	0.4416	0.4115	0.4227	0.3758	0.3534	0.4455
Stage 5	0.3879	0.4179	0.4446	0.3908	0.4191	0.4434
Stage 6	0.4749	0.4913	0.4191	0.5178	0.03895	0.397

Note: Sometimes, it may happen that a signature may not be written properly due to the uneven pen quality or ink quality of a pen. Therefore, to reduce such effects in our proposed forecasting process, we have taken 2 signatures at each situation from each person during 2nd to 6th stages of the testing dataset. Then we have calculated the average of the respective features as given below:

$$Average = \frac{(1st\ sample + 2nd\ sample)}{2}$$

The results of the single exponential smoothing technique with $\alpha = 0.1, 0.2, \ldots\ldots, 0.9$ are given in Table 4.

Table 4. Single exponential smoothing using $\alpha = 0.1, 0.2, \ldots\ldots, 0.9$

Actual		Forecast								
Time		$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$	$\alpha = 0.6$	$\alpha = 0.7$	$\alpha = 0.8$	$\alpha = 0.9$
1	876	-----	-----	-----	-----	-----	-----	-----	-----	-----
2	812	876	876	876	876	876	876	876	876	876
3	754	869.6	863.2	856.8	850.4	844	837	831.2	824.8	818.4
4	768	858.04	841.36	825.96	811.84	799	787.44	777.16	768.16	760.44
5	725	849.0360	826.6880	808.5720	794.3040	783.5	775.7760	770.7480	768.0320	767.2440
6	784	836.6324	806.3504	783.5004	766.5824	754.25	745.3104	738.7244	733.6064	729.2244
7	768	831.3692	801.8803	783.6503	773.5494	769.1250	768.5242	770.4173	773.9213	778.5224
8	-----	825.0323	795.1042	778.9552	771.3296	768.5625	768.2097	768.7252	769.1843	769.0522

Next from Table 4, calculating MAD for $\alpha = 0.1$, we get

$$MAD = \frac{|(-64)+(-115.6000)+(-90.0400)+(-124.0360)+(-52.6324)+(-63.3692)|}{6}$$

$$= 84.9463$$

The remaining values of MAD for different values of smoothing constants are given in Table 5.

Table 5. Comparison of values of MAD.

Values of α	Values of MAD
0.1	84.9463
0.2	67.4131
0.3	54.0803
0.4	49.4185
0.5	45.7292
0.6	42.7383
0.7	40.6335
0.8	39.0512
0.9	40.5837

From Table 5, the smallest value of MAD is found to be 39.0512 for α = 0.8. Hence we have considered the next short-term forecasted value as 769.1843 (highlighted in Table 4).

Now, applying our proposed model, the resultant medium or long-term signature forecasting values are obtained as below:

$$Lower\ Limit = 769.1843 - \frac{769.1843}{4}$$

$$= 576.8882$$

$$Upper\ Limit = 2 \times 576.8882$$

$$= 1153.8$$

Thus the required range is [576.8882, 1153.8].

The above range can be tested with tables from Table 2 and 3.

Proceeding similarly for HWR feature, the required range is obtained as [0.3337, 0.6674] and can be tested from Table 2 and 3.

3.1 Performance measurement

Here, we have evaluated performance measurements of the proposed predictive modeling system in terms of Positive Predictive Values (PPV) and False Rejection Rate (FRR) as given below (Altman, & Bland, 1994):

(i) Area:

Let

X : Number of signature areas falling within the specified range of predictive area.

$\overline{X}$: Number of signature areas that are not within the specified range of predictive area.

We get,

$$X = 396$$

$$\overline{X} = 54$$

$$PPV = \frac{396}{(396+54)} \times 100$$

$$= 88\ \%$$

$$FRR = \frac{54}{(396+54)} \times 100$$

$$= 12\ \%$$

(ii) HWR:

Similarly, Let

Y : Number of signature Height-Width Ratios (HWR) fall within the specified range of predictive HWR.

$\overline{Y}$: Number of signature Height-Width Ratios (HWR) that are not within the specified range of predictive HWR.

We get,

$$Y = 431$$

$$\overline{Y} = 19$$

$$PPV = \frac{431}{(431+19)} \times 100$$

$$= 95.78\ \%$$

$$FRR = \frac{19}{(431+19)} \times 100$$

$$= 4.22\%$$

Note:

1. Since we have taken averages of the features at each situation, at each time and thus reduce the testing dataset from 900 to 450, therefore we can take 450 in place of 900.
2. In the proposed predictive system, we have considered only the genuine signatures, therefore we have evaluated performance measurement in terms of FRR (along with PPV) only, but not with FAR and AER.

4. Conclusions

The proposed system gives fairly good accuracy for Area feature and very good accuracy for HWR feature considering the limitations under which the study was conducted. In the study the signatures forecasting was a tedious process considering both the period of data collection which helps us to prepare historical data as well as the situations of variations. Usually, a predictive modeling system seems to deal with secondary data, but the proposed system has to apply primary data due to lack of official ready-to-use continuous data. Secondly during every stage of data collection, each of the involved participants has to vary their physical situations as well as writing materials and surfaces. Thus the data collection process becomes more time consuming, which is another limitation of the existing system. Here, we have forecasted the signatures only with two features. This may be extended to other features of signatures. Also, with more availability of historical data, we can test proposed system with other technique like the ARIMA modeling technique. As far our knowledge such type of work has not been conducted before and hence a comparative study was not possible. In most of the cases, due to the intrapersonal variations, a genuine signature gets rejected. Here, time factor is one of the main factors for rejecting a genuine signature. Because as the time passes, a person's signature may slightly change which does not mean that specific signature is a forgery. Thus we are required to forecast a genuine signature to observe its changes. Hence after forecasting, our next work will be to verify the forecasted signature such that the system can correctly identify the genuine signatures as "genuine" and forgeries as "forgery". Hence the proposed study has importance in the context of current status in all organizations where a person' s signatures is required to establish the person's identity.

References

Altman, D. G., & Bland, J. M. (1994). Statistics notes: Diagnostic tests 1: sensitivity and specificity. *British Medical Journal, 308*, 1552. doi:10.1136/bmj.308.6943.1552

Azzopardi, G. (2006). *How effective are radial basis function neural networks for offline handwritten signature verification?* (Project report, University of London, England). Retrieved from https://www.researchgate.net/publication/265672969_How_effective_are_Radial_Basis_Function_Neural_Networks_for_Offline_Handwritten_Signature_Verification.

Bertino, A., & Bertino, P. (2011). History of forensic handwriting analysis. In forensic science: Fundamentals and investigations 2012 update, Cengage Learning, (pp: 279-282). Retrieved from http://www.cvisd.org/site/handlers/filedownload.ashx?moduleinstanceid= 5221&dataid= 12155&FileName= Forensic-Science-Fundamentals-Investigations-book.pdf

Bertolini, D., Oliveria, L. S., Justino, E., & Sabourin, R. (2010) Reducing forgeries in writer-independent off-line signature verification through ensemble of classifiers. *Pattern Recognition, 43*, 387-396.

Fotak, T., Baca, M., & Koruga, P. (2011). Handwritten signature identification using basic concepts of graph theory. *Wseas Transactions on Signal Processing, 7*(4), 117-129. Retrieved from http://dl.acm.org/ citation.cfm?id=2190015

Gonzalez, R. C., Woods, R. E., & Eddins S. L. (2010). *Digital image processing using matlab*. New Delhi, India: Tata McGraw Hill Education.

Houck, M. M., & Siegel, J. A. (2015). *Collection of handwriting exemplars. In fundamentals of forensic science*. Cambridge, MA: Academic Press.

Scheidat, T., Vielhauer, C., & Dittman, J. (2009). Handwriting verification-Comparison of a multi-algorithmic and a multi-semantic approach. *Image and Vision Computing, 27*(3), 269-278.

Mahanta, L. B., & Deka, A. (2013). A study on handwritten signature. *International Journal of Computer Applications, 79(2)*, 975–8887.

Jain, C. L., & Malehorn, J. (2005). *Exponential smoothing. In practical guide to business forecasting*. New York, NY: Graceway Publishing.

Karouni, A., Daya, B., & Bahlak, S. (2011). Offline signature recognition using neural networks approach. *Procedia Computer Science, 3*, 155-161. doi:10.1016/j.procs2010.12.027

Kisku, D. R., Gupta, P., & Sing, J. K. (2010). Offline signature identification by fusion of multiple classifiers using statistical learning theory. *International Journal of Security and Its Applications, 4*(3), 35-45. Retrieved from http://www.sersc.org/journals/IJSIA/vol4_no3_2010/3.pdf

Mahanta, L. B., & Deka, A. (2013). Skew and slant angles of handwritten signature. *International Journal of Innovative Research in Computer and Communication Engineering*, *1*(9), 2030-2034. Retrieved from https://www.ijircce.com/upload/2013/november/7_Skew.pdf

Makridakis, S., Wheelwright, S. C., & Hyndman, R. J. (2008). *Forecasting methods and applications* (3rd ed.). Hoboken, NJ: John Wiley and Sons.

Shah, J. (2009). *Supply chain management: Text and cases* (pp. 165-167). Delhi, India: Pearson Education India. Retrieved from https://www.safaribooksonline.com/library/view/supply-chain-management/9788131760994/