ISSN: 2349-6495

Distance Based Models of Keystroke Dynamics User Authentication

Soumen Roy[1], Utpal Roy[2], D. D. Sinha[3]

[1,3]*Department of Computer Science and Engineering, University of Calcutta, Calcutta, India*
2Department of Computer & System Sciences, Visva-Bharati, Santiniketan, India

Abstract**—***Distance based algorithms are used in pattern recognition techniques. This is not new, but in this paper we have implemented 20 different algorithms in R statistical programming language and calculated their performance, so we can compare their performance soundly. We have executed all the algorithms on our own keystroke database, which we have collected from 12 individuals during 4 months.*

Keywords—Keystroke Dynamics, Behavioral biometric, ROC Curve, Manhattan Distance, Euclidean Distance, Mahanobolis Distance, Z Score.

I. INTRODUCTION

Keystroke dynamics is a method of analysing the way a user types on a keyboard and classify the user based on their regular typing rhythm. Here, users are well-known by their typing style much like face prints, finger prints, voice prints, signature etc. It is very economic and cannot be lost or stolen in addition with it can be easily integrated in any existing knowledge-based user authentication with small alternation.

Our typing style can be easily calculated by simple key event program. In our experiment we have implemented Java Applet program to get the raw data of keystroke press and release timing pattern where getTime() function return the time of key press and release events. Then we have calculated the following features of keystroke dynamics: key hold time (KD), up-up key latency (UU), up-down key latency (UD), down-up key latency (DU), down-down key latency (DD), total time (ttime), tri-gap time (trigap) and four-gap time (4gap).

Keystroke Dynamics as biometrics characteristics is not a new one. First time, in the year 1897, Bryan and Harter investigated keystroke dynamics. In 1975, Spillane described the concept of keystroke dynamics and suggested in an IBM technical bulletin that typing rhythms might be used for identifying the user at a computer keyboard. Forsen et al. in 1977 conducted preliminary tests of whether keystroke dynamics could be used to distinguish typists. Gaines et al. in 1980 produced an extensive report of their investigation with seven typists into keystroke dynamics. After then S. Bleha submitted his PhD thesis on Recognition system based on keystroke dynamics in 1988 [1]. R. Joyce and G. Gupta proposed an identity authentication based on keystroke latencies in 1990 [2]. F. Monrose et al. [3] proposed keystroke dynamic as a biometric for authentication in 2000. Different online and offline applications already have been done by fixed text and free text keystroke dynamics. Keystroke dynamics research has been going on for the more than thirty three years. Many methods have been proposed during that time. Methods based on traditional statistics-such as mean times and their standard deviations are common. Over the years, different pattern recognition methods have come into vogue and been applied to keystroke dynamics; neural networks, Fuzzy logic and support vector machines among others. Many classification algorithms have been proposed and many databases are available in the Internet. In evaluation process of different classifiers on different database, we have obtained different average Equal Error Rates (EERs) because selecting the string for the database and considering the features for classification affect the error rate. It has been established that our typing styles are similar for the common daily used words (name, address, e-mail ID etc.). In this connection we have chosen the daily used words to train the system.

We have collected press and release time of 12096 keystrokes of 1440 samples of patterns from 12 different individuals in 4 different sessions with minimum of one month interval for five different common words (“kolkata123”, ”facebook”, ”gmail.com”, ”yahoo.com”, ”123456”) in our experiment. Then we have considered all 8 different features and combination of features then we have executed 8 different classifiers on that collected data. In our observation we got 2.4% of EER for the classifier OutlierCount (z-score) by taking all the features in our consideration. In second position NaïveBaysian classifier given 5.3% of EER when we have taken in our consideration all the features and all 4 strings (“kolkata123”,”facebook”, ”gmail.com”, ”yahoo.com”). So the adaptation of keystroke dynamics technique in any existing system increases the security level upto 94.7% to 96.6%.

II. BACKGROUND DETAILS

In 30+ years of experience, many researchers have proposed their algorithms, taking various features, various length of pattern string.

Table 1.Background of keystroke dynamics

Authors	Classifiers	Length of the pattern	Features	EER (%)
Joyce & Gupta [2]	Manhattan	33	UD	0.25-16.36
Bleha et al. [1]	Euclidian	11-17	UD	2.8-8.1
Haider et al. [7]	Nural Network	7	UD	16.1
Yu & Cho [5]	SVM	6-7	UD	10.2
Killourly S. [4]	Manhattan (Scaled)	10+	UD	9.6
Kang et al. [6]	K mean	7-10	KD, UD	3.8
Giot et al. [8]	SVM	100	KD, UD	15.28

III. EXPERIMENTAL RESULTS

We have implemented a program in Java Applet for collecting, which has the capability of capturing all key pressing and releasing events, which are used to create the database of different sample of passwords and timing templates. Here we have calculated average equal error rate for all eight algorithms considering some single feature and combination of features for all five strings.

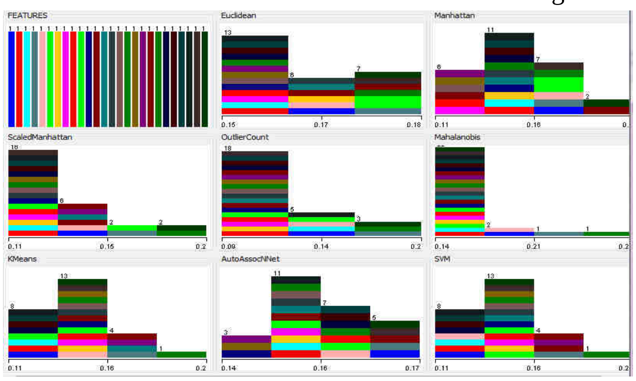


Fig.1. Histogram of the string "kolkata123"

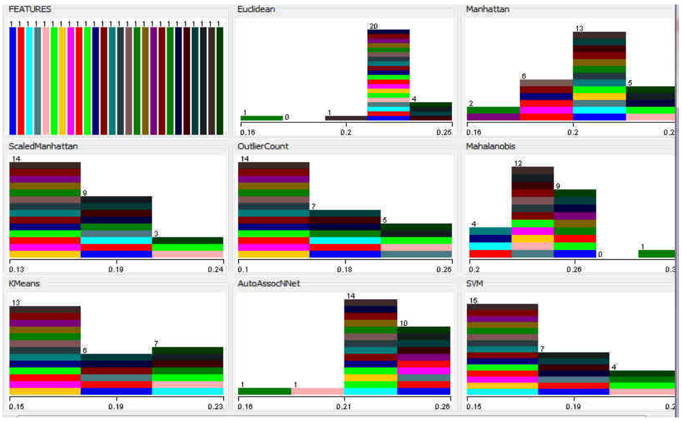


Fig. 2. Histogram of the string "yahoo.com"

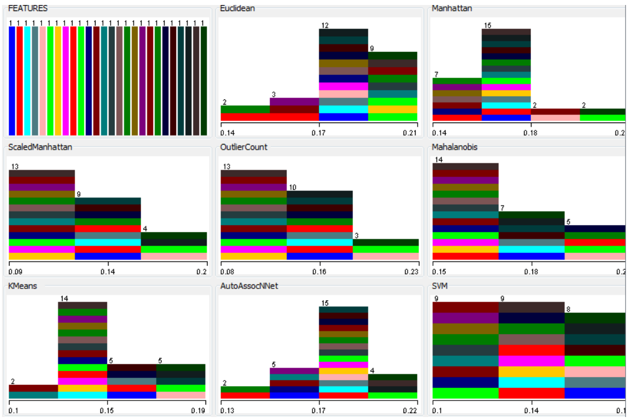


Fig. 3. Histogram of the string "gmail.com"

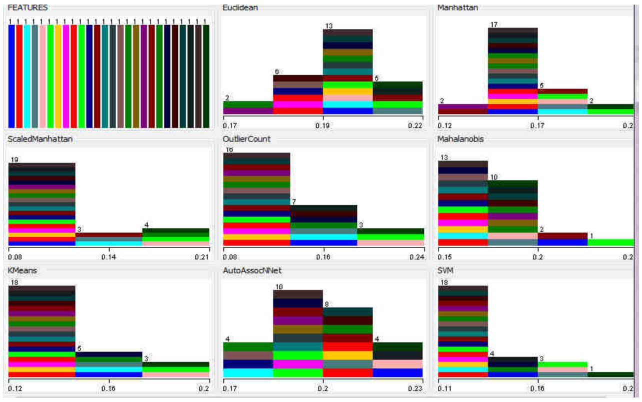


Fig. 4. Histogram of the string "facebook"

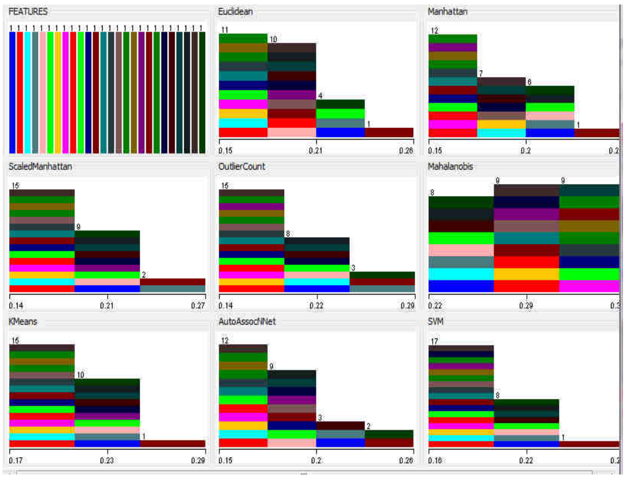


Fig. 5. Histogram of the string "123456"

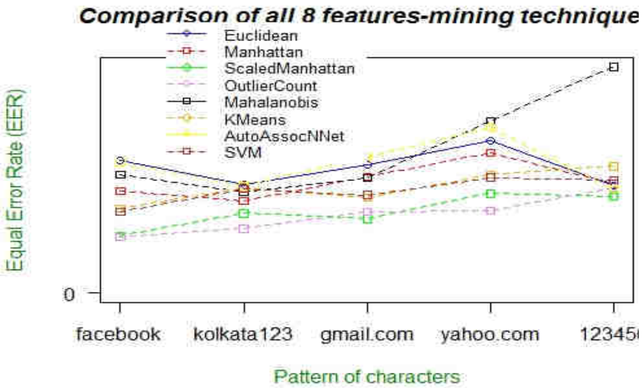


Fig. 6. Line chart of all 8 classifiers

In the above figure, we see that for all the strings ourlierCount (z-score) is achieved best result after scaled Manhattan.

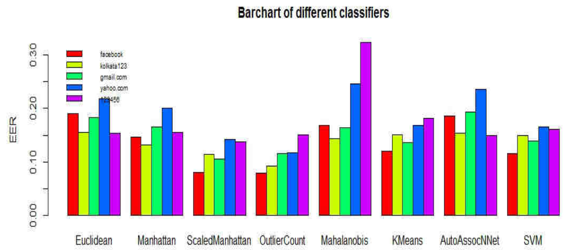
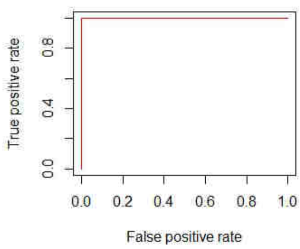
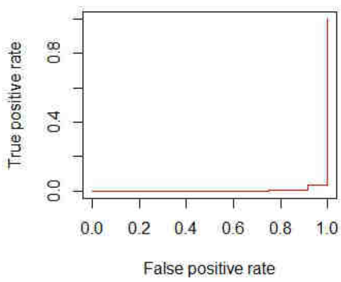


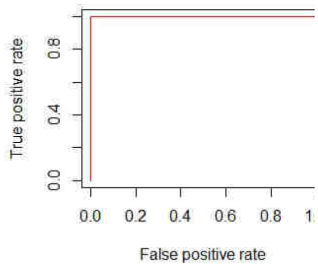
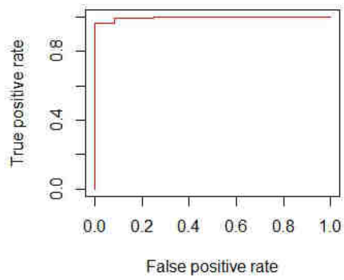
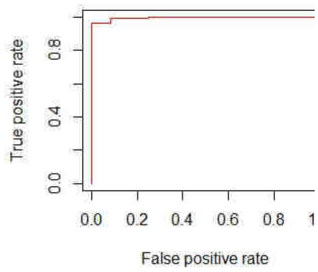
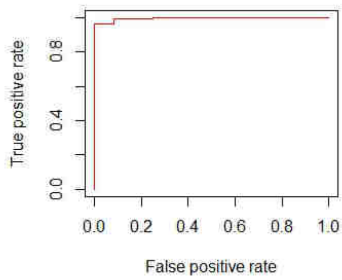
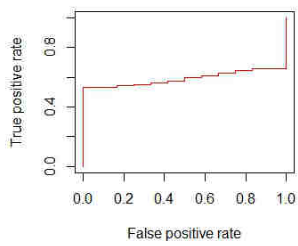
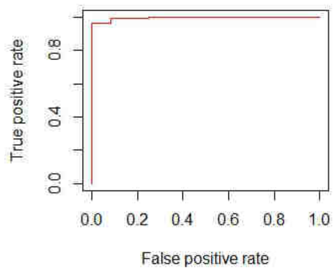
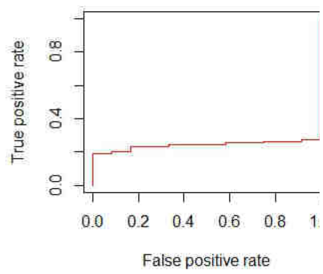
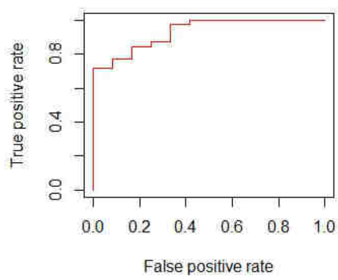
Fig. 7.Bar chart of all 8 classifier

Here we see that no combination of features and algorithms give bellow 0.08 average equal error rate for all five type of fixed string.

We have tested combining these five strings and we got the following result. Here minimum average equal error rate is 0.024 where all five strings and all features are considered.

Table 2.ROC curve of all 2o distance based algorithm

Name of the Algorithms	**ROC**	**Name of the Algorithms**	**ROC**
Chebyshev		Ruzicka	

Canberra		Soergel	
Czekanowski		Sorensen	
Gower		Wavehedges	
Intersection		Euclidean	

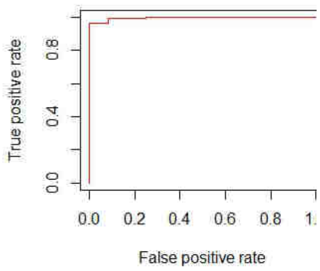
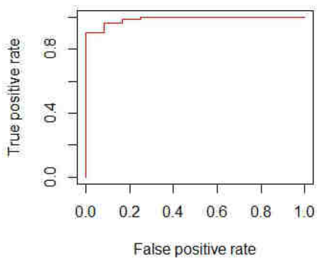
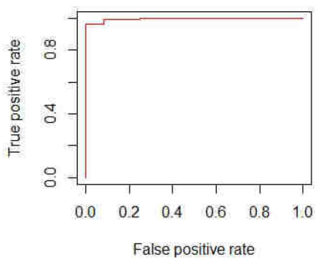
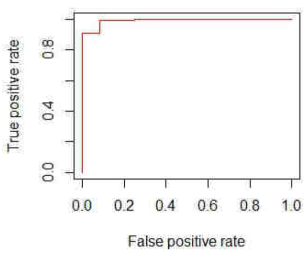
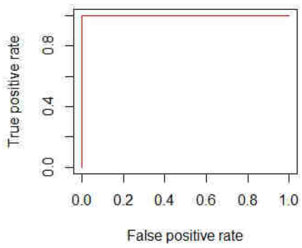
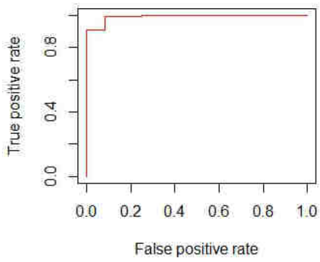
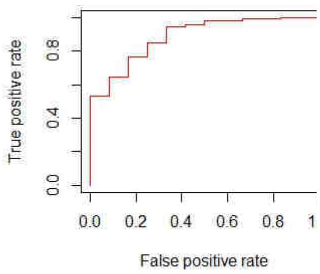
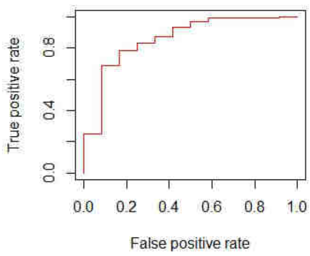
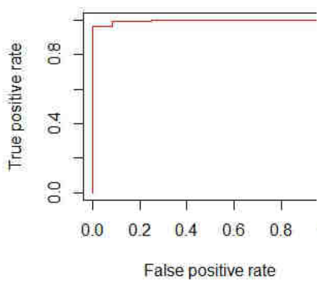
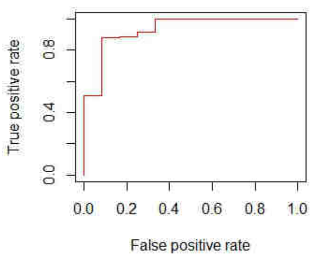
<table>
<tr><td>Kulczynski</td><td>
</td><td>Manhattan</td><td>
</td></tr>
<tr><td>Kulczynskis</td><td>
</td><td>ScaledManhattan</td><td>
</td></tr>
<tr><td>Lorentzian</td><td>
</td><td>OutlierCount</td><td>
</td></tr>
<tr><td>Minkowski</td><td>
</td><td>Mahalanobis</td><td>
</td></tr>
<tr><td>Motyka</td><td>
</td><td>KMeans</td><td>
</td></tr>
</table>

Table 3.EER of all 20 distance based algorithms

Classifiers	**EER**	**Sd**	**Classifiers**	**EER**	**Sd**
Chebyshev	0.083	0.289	Ruzicka	0.871	0.109
Canberra	0.071	0.104	Soergel	0.129	0.109
Czekanowski	0.129	0.109	Sorensen	0.129	0.109
Gower	0.515	0.264	Wavehedges	0.129	0.109
Intersection	0.579	0.255	Euclidean	0.205	0.123
Kulczynski	0.129	0.109	Manhattan	0.144	0.127
Kulczynskis	0.129	0.109	ScaledManhattan	0.088	0.097
Lorentzian	0.044	0.076	**OutlierCount**	**0.024**	**0.072**
Minkowski	0.219	0.119	Mahalanobis	0.260	0.181
Motyka	0.129	0.109	KMeans	0.184	0.095

IV. CONCLUSION

This is the first time we have executed 8 different classification algorithms on 5 similar keystroke database taking in our consideration all 8 features and combination of features so we can compare the classifiers on an equal basis. In our evaluation process, we have identified the best classifier (z-score). It achieved 91.2% of accuracy for the string "kolkata123" (considering KD, DU, UD, Trigap and 4gap timing features), 90.5% of accuracy for the string "yahoo.com" (considering KD, UD), 91.7% of accuracy for the string "gmail.com" (considering KD, UD), 92.0% of accuracy for the string "facebook" (considering KD, DD, UU, DU, UD and Trigap), 85.5% of accuracy for the string "123456" (considering KD, DU, Trigap and 4gap timing features). Z-score classification algorithm gives the highest accuracy for all the string patterns. We also have tested this algorithm on the entire strings database and we got 97.6 % of accuracy. So it has been established that this technique can be used as a safe guard of password or PIN in knowledge-based user authentication. But in practical there are many affecting factors may affect way of this process. Need much more experiment on it like key pressure; finger placement etc. can be calculated.

REFERENCES

[1] Bleha, S. et al. (1990). Computer-access security systems using keystroke dynamics. IEEE Transactions on Pattern Analysis and Machine Intelligence 12, 1217–1222.

[2] Joyce, R. & Gupta, G. (1990). Identity authorization based on keystroke latencies. Communication of ACM 33 (2) 168–176.

[3] Monrose, F. & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. Future Generation Computer Systems, Vol. 16, No. 4, pp. 351–359.

[4] Killourhy, K. S. (2012). A Scientific Understanding of Keystroke Dynamics. PhD thesis, Computer Science Department, Carnegie Mellon University, Pittsburgh, US.

[5] E. Yu and S. Cho, "Novelty detection approach for keystroke dynamics identity verification," in Intelligent Data Engineering and Automated Learning, vol. 2690, pp. 1016–1023, Springer, Berlin, Germany, 2003.

[6] P. Kang, S. S. Hwang, and S. Cho, "Continual retraining of keystroke dynamics based authenticator," in Advances in Biometrics, Proceedings, vol. 4642, pp. 1203–1211, Springer, Berlin, Germany, 2007.

[7] S. Haider, A. Abbas, and A. K. Zaidi, "A multi-technique approach for user identification through keystroke dynamics," in Proceedings of the 2000 IEEE Interantional Conference on Systems, Man and Cybernetics, vol. 2, pp. 1336–1341, October2000.

[8] R.Giot, M. El-Abed, andC. Rosenberger, "GREYCkeystroke: a benchmark for keystroke dynamics biometric systems," in Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems (BTAS '09), pp. 1–6, September2009.

[9] Pin ShenTeh, Andrew Beng Jin Teoh, and ShigangYue, A Survey of Keystroke Dynamics Biometrics, The ScientificWorld JournalVolume 2013, Article ID 408280, 24 pages

[10] S. Roy, U. Roy, D.D. Sinha, "Enhanced Knowledge-Based User Authentication Technique via Keystroke Dynamics", International Journal of Engineering and Science Invention (IJESI), Vol 3, Issue 9, Sep, 2013, 41-48.