UDC 004.896
DOI: 10.20535/SRIT.2308-8893.2022.2.10

RESEARCH OF AUTOENCODER-BASED USER BIOMETRIC VERIFICATION WITH MOTION PATTERNS

M.P. HAVRYLOVYCH, V.Y. DANYLOV

Abstract. In the current research, we continue our previous study regarding motion-based user biometric verification, which consumes sensory data. Sensory-based verification systems empower the continuous authentication narrative – as physiological biometric methods mainly based on photo or video input meet a lot of difficulties in implementation. The research aims to analyze how various components of sensor data from an accelerometer affect and contribute to defining the process of unique person motion patterns and understanding how it may express the human behavioral patterns with different activity types. The study used the recurrent long-short-term-memory autoencoder as a baseline model. The choice of model was based on our previous research. The research results have shown that various data components contribute differently to the verification process depending on the type of activity. However, we conclude that a single sensor data source may not be enough for a robust authentication system. The multimodal authentication system should be proposed to utilize and aggregate the input streams from multiple sensors as further research.

Keywords: motion patterns recognition, biometric verification, recurrent autoencoders.

INTRODUCTION

In the modern world, we strive to automate all the systems and processes but to keep stability, reliability, and trustworthiness of automatization on the highest possible level. Although there are still a lot of use cases, which are impossible to make fully automated due to ethical and other artificial intelligence issues (for example, in the medical area), automated ensuring of the system reliability and stability is critical in many real-world applications. There is a need for advanced monitoring and anomaly detection applications for this purpose. The appliance of such systems is very broad: financial operations, sensors or medical data, security sector, and many others.

All of such systems meet some challenges during development as there is a lot of uncertainty. First, you never know what anomaly data should look like, despite some apparent cases, especially if you have an inlier type of anomaly. Second, all the processes change over time, so the data – and if new data points are not similar to the previous ones – it does not mean they are anomalies. As well, real-world data frequently has a pretty complex structure and is heterogeneous, leading to the inefficiency of classic statistical and machine learning methods.

From the machine learning perspective, it is impossible to apply a classical supervised approach, as in most cases, only regular data examples are available. In this study, we will look at the user motion-based verification problem. In this application, it is not feasible to solve such a problem in a classic supervised manner due to the huge population and inability to compare specific users with all others (as well, there could be personal information concerns). Therefore, only self-supervised or unsupervised approaches may fit for problem solutions.

The research object is motion-based user biometric verification based on sensor data.

The purpose is to research and analyze various components, which affect the result of verification and understanding how sensor data (accelerometer) and its components may express human behavioral patterns with different activity types.

LITERATURE REVIEW

According to [1], biometric data can be physiological: like iris, face, fingerprint, and behavioral: like motion or gate patterns, mouse or keyboard movements, or even both (voice data). For example, in [2], the authors propose interesting approaches for biometric identification systems based on circular kernel principal component analysis, Chebyshev transforms hashing, and Bose–Chaudhuri–Hocquenghem error-correcting codes for ear-based biometrics.

Nevertheless, many biometric-based verification systems are explicit and require specific actions by the person or from an external supervisor. Many biometric approaches utilize computer vision approaches like iris, ear, face, or gate-based, which require a camera sensor. A camera sensor is not convenient for continuous authentication, and for explicit authentication, somebody, for example, should take a photo. Usage of the camera for continuous and implicit authentication may cause problems: it should process a considerable amount of data (video or image streaming), there are issues with saving and processing personal data, and a person may feel uncomfortable as they are constantly watched. On the other hand, the motion-based or mouse/keyboard movement authentication, that can be implicit and be conducted all the time in the background without user intervention, such data is more lightweight, and in most cases do not need additional sensor installation, because most devices already have all the sensors for other purposes.

We can reformulate the motion-based user verification task as a time series classification problem. The anomaly detection on time series data has already been a research topic in many fields. As well, time-series data anomalies, according to [3], have their taxonomy: point outliers, subsequent, and time series. In user verification, the subsequent outliers are of the most significant interest, as we can reformulate the task from a pattern recognition and classification point of view. Worth mentioning that the time series classification task has a lot of applications, besides anomaly detection, especially in the field of motion sensory data.

In [4], we can see the survey on existing machine learning algorithms, both statistical and deep learning, for anomaly detection on univariate time series. In the case of univariate data – deep learning methods do not overperform classic machine learning and statistical methods in the survey. However, for multivariate and heterogeneous time series anomaly detection, according to [5], deep learning methods are the ones with the greatest accuracy. In [6], we can review the comparative study for anomaly detection on multivariate time series for LSTM and CNN-based (Temporal Convolutional Network) architectures, and the CNN-based solution even slightly outperformed the recurrent one. In [7], the CNN-based and LSTM model are also compared for biometric motion-based verification on various datasets and activity types. The hybrid LSTM-CNN-based neural network was proposed, as it allows to capture of temporal dependencies on more complex features extracted with convolutional layers.

In [4], the authors measures not only predictions quality metrics but the computational time for model training and inference and conclude that amongst all deep learning approaches, the best combination of quality metrics and computational time has an autoencoder-based approach, which is very important in the angle of verification and security to achieve best results with the lowest latency.

In our previous research [8], we conducted a comparative analysis of various types of recurrent autoencoders for user biometric verification. We compared them to classical machine learning algorithms such as Isolation Forest and One-Class SVM. In [9], the authors designed the user personalization and biometric verification system based on geometric concepts on the convex hull. They mentioned the SVM as a state-of-the-art approach (in 2012), but the one which requires a lot of time for training, thus do not fit for on-the-fly verification model training. In our research, One-Class SVM showed the lowest and most unstable performance, though. In contrast, the variational autoencoders showed the best performance; even so, all types of autoencoders provide comparable results.

It is essential to mention that despite the continuous authentication idea gaining popularity, some research shows that it still underperforms compared to explicit biometric systems. In the [10], authors mention that the multimodel authentication systems are a better choice. It allows achieving the highest level of system robustness, increasing flexibility, and mitigating drawbacks of every verification modality. In [11], authors propose multimodal authentication based on face, motion patterns, and touch stroke. Motion-based authentication systems can also be boosted using multiple sensors [10, 13], such as accelerometers and gyroscopes. The combination of multiple sensory data increases the accuracy of the motion-based authentication system.

It is relevant to note that the deep learning model, especially autoencoder-based, may be used as a single base model for many purposes and may have interesting applications – for example, activity classification, and more exotic things like detecting smoking events [12] or gender of the person [13]. Overall, having a single model for a couple of purposes is a tendency in the modern machine learning and data science field. It allows to reduce the cost and add additional context for every task-specific downstream model appliance at the same time.

In this study, we wanted to continue the in-depth analysis of biometric motion-based user verification and conduct detailed experiments and research regarding how accelerometer motion data and its components describe the person using recurrent autoencoders.

MATERIALS AND METHODS

Based on our previous research [8], we will use the undercomplete autoencoder as a baseline model architecture with recurrent LSTM layers types, as all types of autoencoders (undercomplete, variational, contractive) showed comparable results between each other.

Autoencoder architecture consists of an encoder and decoder. Autoencoders are learned to reconstruct the input data points from some hidden space. Therefore, the optimization task objective is to minimize reconstruction error:

$$E = \sqrt{\sum_{i=1}^{n} \left\| x_i - d_{\varphi}(e_{\theta}(x_i)) \right\|}\,, \tag{1}$$

where $x_1, \ldots, x_n$ is data rows, d is the decoder, and e is the encoder with some parameters.

In the undercomplete autoencoder type, the data is encoded in lower-dimensional space. Such compression guarantees that the model will not blindly memorize the train set but will learn the proper feature space and data embedding, which later as well can be used for various purposes. The decoder's purpose is to recreate the sample from an encoded example.

The optimization process of encoder and decoder parameters is done with classic backpropagation using gradient descent-based algorithms to minimize (1) [14].

After model training, the decision threshold ε should be set. This threshold will be used at the model inference step: the data points that reconstruction error will be higher than this threshold would be considered anomaly or non-self class, and self otherwise. The threshold setting process is dependent on the use case and varies for different applications. It can be set manually by an expert or some knowledge keeper or automatically based on the error distribution on some predefined dataset.

We will use an autoencoder with LSTM layers. The LSTM layer contains cells with specific internal structures. The memory cell contains three gates: input, output, and forget. The input controls the input activations when the output controls the output flow. The forget gate controls what information memory cells should forget and what to pass further through the network, which theoretically may solve the problems of long sequences, which is a known problem for vanilla recurrent neural networks, which fail to learn from big sequences [15].

EXPERIMENTS

Dataset: open-source dataset [16] with accelerometer data (52 Hz) from 15 people with seven activity types. In [9], the original paper that presents the used dataset, the HAR (human activity recognition) task should be solved first, as training the algorithm on all types of activities will mostly bring additional noise to data and decrease the metrics due to not enough amount of data points per each activity. The problem can be understood as detecting the unique and distinguishing patterns of a specific person's motions.

For understanding accelerometer data and human movement patterns, we will train the model separately for each axis (*x*, *y*, *z*), axis pairs (*xy*, *yz*, *xz*), and all three axes — *xyz* for different types of activities.

For deep learning models, we split data in overlapping on 50 percent windows with a length of 52 (because of accelerometer frequency).

We split the original dataset into a 33% share for the test set and the rest for the train.

An accelerometer measures changes in velocity along one axis (Fig.1). The values reported by the accelerometers are measured in increments of the gravitational acceleration, with the value 1,0 representing an acceleration of 9,8 meters per second (per second) in the given direction. Depending on the direction of the acceleration, the sensor values may be positive or negative [17].

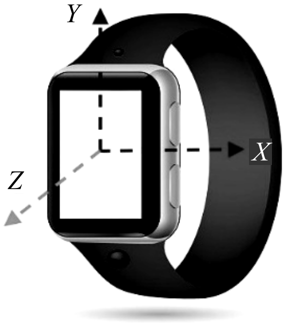

Fig. 1. Accelerometer axis demonstration on the general smartwatch/bracelet [12]

Autoencoder model architecture is in Table 1 below. As an activation function, the hyperbolic tangent was used. The model was trained in 10 epochs, with Adam optimizer and mean absolute error loss. We build autoencoders with python on Keras library and Tensorflow backend [17].

The dropout rate was 0,4.

Table 1. The used model architecture with layers' type, shape, and amount of params

Type	Layer Shape	# of Params
LSTM	(None, 52, 20)	1760
LSTM	(None, 10)	1240
RepeatVector	(None, 52, 10)	0
Dropout	(None, 52, 10)	0
LSTM	(None, 52, 10)	840
LSTM	(None, 52, 20)	2480
TimeDistributed	(None, 52, 3)	21

The threshold formula was used as in [8]:

$$T = \sum_{i=1}^{N} {MAE_i}/{n} + std(MAE_i)\,, \qquad ,$$

where *MAE* is the mean absolute error between ground truth and predicted sample; *std* – standard deviation, and *N* is the number of samples in the training dataset.

As model evaluation metrics [19], the *EER* (equal error rate), *FAR* (false accept rate) and *FRR* (false reject rate), and ROCAUC (area under the curve) were chosen, which are typical for assessing the biometric system quality:

$$FAR = FPR = \frac{FP}{FP + TN}\,; \quad FRR = FNR = \frac{FN}{TP + FN}\,.$$

Equal error rate (*EER*) (illustrated in Fig. 2) is obtained by adjusting the system's detection threshold to equalize *FAR* and *FRR*. The *EER* is calculated using the following formula:

$$EER = \frac{FAR + FRR}{2}\,,$$

where $|\,FAR + FRR\,|$ is the smallest value [20].

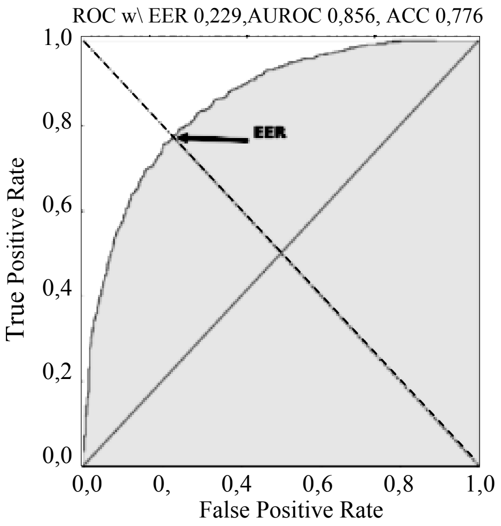

Fig. 2. Illustration of the *EER* calculation from [19]

RESULTS

For analysis, we considered only 1, 3, 4, and 7 types of activities. We filtered out 2,5, and 6 activities (Standing Up, Walking and Going updown stairs, Going UpDown Stairs, Walking and Talking with Someone) as less than 200 data points were presented in the train set and less than 100 in the median for 15 users (Table 2).

Table 2. Amount of samples in train set (median value for 15 users) for various activities

Activity	1	2	3	4	5	6	7
# in trainset (median for 15 users)	133	89	301	599	86	62	860

We can review the results for every metric in the tables below: *EER* and *AUC* (Table 3), *FAR*, and *FRR* (Table 4). We have 15 users in the dataset therefore we report the average performance metric.

Table 3. Average *EER* and *AUC* (on 15 users) for various activities

Axis data for train	1 activity	7 activity	3 activity	4 activity	1 activity	7 activity	3 activity	4 activity
	Mean *EER*				Mean *AUC*			
x-axis	0,202	0,350	0,407	0,318	0,837	0,671	0,617	0,714
y-axis	0,168	0,276	0,364	0,345	0,868	0,756	0,657	0,687
z-axis	0,161	0,358	0,394	0,237	0,873	0,669	0,631	0,791
x-axis_*y*-axis	0,084	0,200	0,346	0,280	0,938	0,833	0,680	0,759
x-axis_*z*-axis	0,101	0,254	0,358	0,191	0,924	0,777	0,673	0,841
y-axis_*z*-axis	0,081	0,223	0,349	0,225	0,938	0,807	0,680	0,807
x-axis_ *y*-axis_*z*-axis	0,074	0,189	0,334	0,197	0,949	0,846	0,689	0,830

Table 4. Average *FAR* (on 15 users) for various activities

Axis data for train	**1 activity**	**7 activity**	**3 activity**	**4 activity**	**1 activity**	**7 activity**	**3 activity**	**4 activity**
	Mean *FAR*				**Mean *FRR***			
x-axis	0,260	0,527	0,589	0,459	0,067	0,132	0,177	0,113
y-axis	0,207	0,388	0,526	0,517	0,058	0,101	0,160	0,108
z-axis	0,182	0,534	0,564	0,307	0,071	0,128	0,174	0,112
x-axis_y-axis	0,066	0,219	0,464	0,374	0,058	0,116	0,176	0,107
x-axis_z-axis	0,076	0,312	0,488	0,196	0,075	0,133	0,166	0,122
y-axis_z-axis	0,059	0,269	0,459	0,259	0,064	0,117	0,181	0,127
x-axis_ y-axis_z-axis	0,030	0,188	0,452	0,222	0,073	0,120	0,171	0,117

DISCUSSION

The obtained results show that all axes contribute to the final results and hold important information about human patterns. Therefore, training on all three axes shows the best performance.

Another interesting thing we can notice is that for different types of activities – different axis brings more value. For example, in walking (4-th activity), we can say that the model trained on the z-axis only or on another axis in combination with the z-axis has the best performance if looking at *EER* and *AUC*. On the other side, for the 7th activity (Talking While Standing), the y-axis brings the most value.

As well, if looking at the performance of models trained on the x-axis only, they always have the lowest performance compared to others. Still, in combination with the y-axis, the performance increases a lot.

The best results were shown for 1 activity (Working at Computer), but this can be related, that this type of activity has the highest amount of training samples.

Overall, the performance is highly correlated with the number of training samples (Table 1), which may point out for need in artificial synthetic data generation for model training, because for the personalization system is important to be able to work and be reliable as fast as possible. The continuous training of the model during the system is alive should be considered to prevent model and data drift and allow the system to prevent the cold start when there are not that many available training samples.

CONCLUSIONS

In this research, we conducted an in-depth analysis of different components on human motion patterns from sensory data (accelerometer in our case) and whether we can extract distinguishing person patterns from such data and use it for biometric verification systems.

The deep learning approaches have already proved their applicability and stable performance in such cases. Still, as already mentioned, the motion-based authentication shows lower accuracy than other biometric verification (e.g., physiological), but this does not mean that motion-based verification should not be used. The solution to overcome problems drawbacks of various types of verification – is a multimodal authentication system, which increases stability, robustness, and flexibility for customization in different environments.

Looking at the metrics for different accelerometer data components and activities, we can see that every axis contributes to the final result not equally. Depending on the activity type, different features are important, proving that we probably need a multi-stage system with preliminary human activity classification in case of motion-based verification. The advantage of the autoencoder model is that single model can be used for both tasks without the need to train different models.

Further research should be done to create a sensory-based authentication system that utilizes multiple sensors. Such approach should increase the quality of the system but keep the continuous option.

Additional analysis of evaluation metrics should also be done, as there are raising concerns regarding commonly used metrics and biometric evaluation framework, as they may lead to incorrect decisions and be misleading.

REFERENCES

1. S. Minaee, A. Abdolrashid, H. Su, M. Bennamoun, and D. Zhang, *Biometrics Recognition Using Deep Learning: A Survey*. 2021. [Online]. Available: https://arxiv.org/abs/1912.00271. (Accessed on: Feb 15, 2022).
2. L. Olanrewaju, O. Oyebiyi, S. Misra, R. Maskeliunas, and R. Damasevicius, “Secure ear biometrics using circular kernel principal component analysis, Chebyshev transform hashing and Bose–Chaudhuri–Hocquenghem error-correcting codes”, *Signal, Image and Video Processing*, vol. 14, no. 5, pp. 847–855, 2020. doi: 10.1007/s11760-019-01609-y.
3. A. Blázquez-García, A. Conde, U. Mori, and J.A. Lozano, “A Review on Outlier/Anomaly Detection in Time Series Data”, *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–33, 2021. doi: 10.1145/3444690.
4. M. Braei and S. Wagner, *Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art*. 2020. [Online]. Available: https://arxiv.org/abs/2004.00433. (Accessed on: Feb 15, 2022).
5. M. Braei, “Anomaly detection of time series: A comparison of statistical vs classical machine learning vs deep learning approaches”, *Unpublished*, 2019. doi: 10.13140/RG.2.2.17687.80801.
6. S. Gopali, F. Abri, S. Siami-Namini, and A. Namin, “A Comparative Study of Detecting Anomalies in Time Series Data Using LSTM and TCN Models”, *arXiv.org*, 2021. [Online]. Available: https://arxiv.org/abs/2112.09293. (Accessed on Feb 15, 2022).
7. S. Mekruksavanich and A. Jitpattanakul, “Deep Learning Approaches for Continuous Authentication Based on Activity Patterns Using Mobile Sensing”, *Sensors*, vol. 21, no. 22, pp. 7519, 2021. doi: 10.3390/s21227519.
8. M. Havrylovych, V. Danylov, and A. Gozhyi, “Comparative Analysis of using Recurrent Autoencoders for User Biometric Verification with Wearable Accelerometer”, *Proceedings of the 9th International Conference “Information Control Systems & Technologies”*, pp. 358–370, Sept. 2020.
9. P. Casale, O. Pujol, and P. Radeva, “Personalization and user verification in wearable systems using biometric walking patterns”, *Personal and Ubiquitous Computing*, vol. 16, no. 5, pp. 563–580, 2011. doi: 10.1007/s00779-011-0415-z.
10. M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen, “Sensor-Based Continuous Authentication of Smartphones’ Users Using Behavioral Biometrics: A Contemporary Survey”, *IEEE Internet of Things Journal*, vol. 8, no. 1, pp. 65–84, 2021. doi: 10.1109/jiot.2020.3020076.
11. Z. Akhtar, A. Buriro, B. Crispo, and T.H. Falk, “Multimodal smartphone user authentication using touchstroke, phone-movement and face patterns”, *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017. doi: 10.1109/globalsip.2017.8309185.

12. C.A. Cole, D. Anshari, V. Lambert, J.F. Thrasher, and H. Valafar, “Detecting Smoking Events Using Accelerometer Data Collected Via Smartwatch Technology: Validation Study”, *JMIR mHealth and uHealth*, vol. 5, no. 12, pp. e189, 2017. doi: 10.2196/mhealth.9035.
13. A. Mostafa, T. Barghash, A. Assaf, and W. Gomaa, “Multi-sensor Gait Analysis for Gender Recognition”, *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics*, 2020. doi: 10.5220/0009792006290636.
14. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning Adaptive Computation and Machine Learning series*. 2016. 775 p.
15. H. Sak, A. Senior, and F. Beaufays, “Long short-term memory recurrent neural network architectures for large scale acoustic modeling”, *Interspeech 2014*, 2014. doi: 10.21437/interspeech.2014-80.
16. *UCI machine learning repository*. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer. (Accessed on: Feb 15, 2022).
17. F. Chollet, *Building Autoencoders in Keras*. 2016. [Online]. Available: https://blog.keras.io/building-autoencoders-in-keras.html. (Accessed on: Feb 15, 2022).
18. *Getting Raw Accelerometer Events*. [Online]. Available: https://developer.apple.com/documentation/coremotion/getting_raw_accelerometer_events\
19. S. Sugrim, C. Liu, M. McLean, and J. Lindqvist, “Robust Performance Metrics for Authentication Systems”, *Proceedings 2019 Network and Distributed System Security Symposium*, 2019. doi: 10.14722/ndss.2019.23351.
20. Y. Hong and R. Kumar, *Performance Evaluation Metrics for Biometrics-based Authentication Systems*. 2021. [Online]. Available: http://hdl.handle.net/10066/23535. (Accessed on: Feb 15, 2022).

Received 16.02.2022

INFORMATION ON THE ARTICLE

Mariia P. Havrylovych, ORCID: 0000-0002-9797-2863, Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: mariia.havrylovych@gmail.com

Valeriy Ya. Danylov, ORCID: 0000-0003-3389-3661, Institute for Applied System Analysis of the National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Ukraine, e-mail: danilov1950@ukr.net

ДОСЛІДЖЕННЯ БІОМЕТРИЧНОЇ ВЕРИФІКАЦІЇ КОРИСТУВАЧА НА ОСНОВІ АВТОКОДЕРІВ З ПЕРЕВІРКАМИ РУХУ / М.П. Гаврилович, В.Я. Данилов

Анотація. Продовжено попереднє дослідження щодо біометричної перевірки користувача на основі руху з використанням сенсорних даних. Системи сенсорної верифікації розширюють можливості неперервної автентифікації, оскільки фізіологічні біометричні методи, в основному засновані на фото- або відеоданих, стикаються з багатьма труднощами під час реалізації. Мета дослідження — проаналізувати як різні компоненти сенсорних даних від акселерометра впливають і сприяють визначенню процесу унікальних моделей руху людини та розуміння того, як вони можуть виражати моделі поведінки людини з різними видами активності. Як базову модель використано рекурентний автокодувальник довгої-короткої пам’яті. Вибір моделі ґрунтується на попередніх дослідженнях. Результати дослідження показали, що залежно від виду діяльності різноманітні компоненти даних мають різний внесок. Зроблено висновок, що одного джерела даних датчика може бути недостатньо для надійної системи автентифікації. Для подальших досліджень слід запропонувати мультимодальну систему автентифікації, яка повинна використовувати та об’єднувати вхідні потоки від кількох датчиків.

Ключові слова: розпізнавання образів руху, біометрична верифікація, рекурентні автокодувальники.