VOL. 15, NO. 17, SEPTEMBER 2020 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

WINDOW SIZE THRESHOLD ANALYSIS FOR BRAINPRINT IDENTIFICATION USING INCREMENTAL K-NEAREST NEIGHBOUR (KNN)

Siaw-Hong Liew[1], Yun-Huoy Choo[1], Yin Fen Low[2] and Shin Horng Chong[3]
[1]Computational Intelligence and Technologies (CIT) Research Group, Faculty of Information and Communication Technology, Malaysia
[2]Machine Learning and Signal Processing (MLSP) Research Group, Center for Telecommunication Research and Innovation (CeTRI) Faculty of Electronics and Computer Engineering, Malaysia
[3]Robotics and Industrial Automation (CIA) Research Group, Faculty of Electrical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal, Melaka, Malaysia
E-Mail: huoy@utem.edu.my

ABSTRACT

This paper aims to investigate the window size threshold of the incremental update strategy in K-Nearest Neighbours (KNN) for brainprint identification. Electroencephalogram (EEG) signals are low signal-to-noise ratio and non-stationary. Incremental learning is good in handling dynamic applications. It does not require complete training examples; instead it is able to adapt dynamic changes to gradually form the target concept. KNN implements First-In-First-Out (FIFO) strategy to guide the incremental learning updates. The FIFO strategy tends to construct the target concept from the training objects according to availability orders. If the number of training objects exceeds the predefined window size threshold, then the FIFO strategy remove the earliest available object. The step size of training pool is linear increased by 10%, from 20% up to 90%. The classification results showed improvement when the window size threshold is increasing. The optimum results recorded at the window size threshold of 60%, with 0.875 in accuracy, 0.887 in precision and 0.878 in f-measure. The degradation of the classification performance after 60% showed the FIFO incremental update strategy is less promising. Thus, future work should focus on the incremental update strategy for selecting the representative and distinct objects to improve the performance of brainprint identification.

Keywords: k-nearest neighbour (KNN), first-in-first-out (FIFO), brainprint identification.

1. INTRODUCTION

Brainprint identification used the Electroencephalogram (EEG) signals to identify a person from a group of people that are being assessed, which is one-to-N matching. The identification performance is depending on the number of subjects to be classified. In recent years, brainprint identification [1]-[5] or authentication [6]-[8] is growing rapidly due to the outstanding benefits as compared to other biometric modalities (i.e. fingerprint, face, iris, etc.). EEG signals demonstrated the uniqueness among individuals and the aliveness.

One of the well-known classifiers used in pattern recognition is K-Nearest Neighbour (KNN). It is an instance-based learning and it can also be used for incremental learning. KNN had been used in [9] for EEG-based biometric identification. The accuracy achieved 100% for the identification rate when using 64 electrodes. However, it is not applicable in the real-world situations due to the inconvenience caused on the large number of electrodes. Besides that, EEG signals classification through the pattern recognition approach were discussed in [10]. Classifiers such as KNN, Multi-Layer Perceptron (MLP) and Support Vector Machine (SVM) were used to classify the EEG signals recorded in mental letter composing task and mental multiplication task. The MLP is outperformed than the SVM classifiers with the accuracy of 89.17% and 86.67% respectively. The MLP classifier was then compared with the KNN classifier. The accuracy of KNN classifier achieved 93.33% while MLP classifier only yielded 92.50%. Hence, the KNN classifier achieved the best performance as compared to MLP and SVM classifiers.

However, the main challenges faced by EEG signals classification is the non-stationarity over the time and low signal-to-noise ratio. Incremental learning plays a crucial role in dynamic applications[11]. It does not need a full training set, instead it is capable on updating from time to time to include and adapt to the new variations of the target model. The incremental update methods were introduced to cope with the non-stationarity of the EEG signals in order to trace the variations in EEG signals over time [12]. It has been applied in the brain print identification [13] or authentication [14], [15]. The use of the incremental learning managed to improve the classification performance.

KNN algorithm uses First-In-First-Out (FIFO) for incremental update strategy. All the new objects will be included in the training pool if and only if the KNN implements incremental learning. Nevertheless, there is not necessary to include all the new objects as not all the new objects provide representative information. The non-representative objects might degrade the classification performance. Besides, included all the new objects will lead to larger training pool and hence increased the computational complexity.

This paper is presented in few sections as follows: In Section 2 illustrates the experimental materials and methods, which includes data acquisition, data pre-processing and data preparation, feature extraction and

feature selection. Meanwhile in Section 3 demonstrates the KNN algorithm on the incremental update strategy and the window size threshold. In addition, Section 4 portrays the data results, observation with discussion. Section 5 concludes the overall works in conclusion term and suggests the recommendation direction for the future work.

2. MATERIALS AND METHODS

2.1 Data Acquisition

EEG signals was collected from 10 healthy subjects, which consists of 7 males and 3 females. Every subject is in good health condition with normal or corrected normal vision. Informed written consent was signed from all volunteered subjects before the experiment was conducted. The ethical approval and the experimental design have been permitted by the Medical Research and Ethics Committee (MREC) from Ministry of Health Malaysia.

In order to understand the experimental procedures, each subject is requested to read the participant information sheet and requested to sign the consent form before continuing with the EEG signals recording. The subject was sitting on a reclining armchair to give full comfort to avoid the possible artifacts during the recording session. The EEG recording is conducted in a quiet environment.

All the visual stimuli run with the resolution of 700 x 525 pixels and white background at the center of the computer screen was displayed. The display duration for black and white picture is 1 second and followed by 1.5 seconds of white blank screen. Each subject was required with 150 trials, which is 60 trials for selected password picture and 90 trials from the random pictures excluding the password picture. The subject was asked to recognize the chosen password picture and instantly press the mouse when the chosen password picture is displayed on the screen.

All the EEG data were taken and recorded from 21 common active electrodes with the location of FPZ, FP1, FP2, F7, F3, FZ, F4, F8, T3, C3, C4, CZ, T4, OZ, T5, P3, PZ, P4, T6, O2, O1 by using Twente Medical Systems International (TMSi) porti system. The electrodes were used based on the positioning of the International 10-20 electrodes. All the scalp electrodes in the experiment were referred to right earlobe and grounded on right hand of the subject. The sampling rate was set at 512Hz. However, only 5 electrodes (O1, O2, OZ, T5 and T6) were selected and used in this experiment.

2.2 Data Pre-Processing and Data Preparation

Data pre-processing is a necessary and vital step before further analysis is performed. The raw data were pre-processed, i.e. filtered, segmented and rejected the artifacts. The purpose of filtering is to improve the signal quality by minimizing the background noise or interference. A Finite-duration Impulse Response (FIR) filter was used with cut off frequencies of 8-12Hzto obtain alpha band signals. Besides that, the trials with amplitude greater than 100μV were removed during the artifact rejection process.

2.3 Feature Extraction and Feature Selection

Feature extraction is the next step to retrieve the informative EEG signals. In this study, coherence, Power Spectral Density (PSD) and Wavelet Phase Stability (WPS) were used in the feature extraction process. Coherence is used to measure the degree of linear correlation between 2 signals [16] while PSD extracts the correlation information between the measured signals from multiple electrodes channels [17]. WPS quantifies the phase information by using the wavelet-based measure. Phase in signal processing provides more meaningful information than the amplitude [18]. Meanwhile, the Correlation-based Feature Selection (CFS) method was used to reduce the size of feature vectors without degrade the classification performance. CFS is a simple and correlated-based filter algorithm which can applied to discrete and continuous problems [19]. The CFS algorithm assesses the subset of the feature based on the correlation-based heuristic merit. It determines the feature's usefulness through the inter-correlation between the features.

3. K-NEAREST NEIGHBOUR (KNN)

K-Nearest Neighbour (KNN) is a well-known instance-based learning technique that can be found in Waikato Environment for Knowledge Analysis (WEKA) data mining tool and known as instance-based learning with parameter K (IBK) classifier. KNN is a simple and easy-to-use supervised learning algorithm. It can be run in incremental learning by using the knowledge flow interface.

In KNN algorithm, K is the number of nearest neighbours and it is the core deciding factor. The value of K is normally an odd number. The prediction is determined based on the majority class of the nearest neighbours. KNN used the Euclidean distance to calculate the distance between the train and test objects. If the value of K is 1, then the algorithm is the simplest and known as the nearest neighbour algorithm. If the nearest distance of the training object is found, its class will be estimated on the basis of the Euclidean distance for the test object. The formula of Euclidean distance is expressed as follows:

$$\text{distance}(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2}$$

where $X_1 = (x_{11}, x_{12}, \ldots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \ldots, x_{2n})$.

KNN algorithm implements First-In-First-Out (FIFO) for the incremental update strategy. The training pool will include the object whenever the new object is available. The amount of time spent to identify the test object increases linearly with the number of objects in the training pool. Thus, the number of objects in the training pool can be restricted by specifying the window size

threshold. The maximum number of objects allowed in the training pool is specified by the window size threshold. The inclusion of new objects greater than the value of the window size would result in removing the old objects.

Window size threshold is required if and only if the user would like to control the size of training pool. Nevertheless, the determination of the window size threshold is an essential process in order to avoid the uncertainty knowledge in the training pool. Hence, an investigation on the window size threshold is carried out to determine the appropriate window size threshold for the brainprint identification model.

In this study, only 10% is used for the training data while 90% is used for testing data. The 10% training data is equivalent to 148 training objects. Hence, the window size is started with 296, which is 20% from the total data, and the window size threshold increased up to 90% with the step size of 10%.

4. RESULTS AND DISCUSSIONS

In this section, the classification performance for brainprint identification is discussed. It is assessed based on the accuracy, precision and f-measure. Initially, the data was split into 10% for the training data while90% for the testing data. Firstly, the brainprint authentication model is tested without implements the incremental learning and the classification performance is demonstrated in Table-1.

Table-1. Brainprint identification performance for 10 Subjects.

Subject	Accuracy	Precision	F-Measure
1	0.731	0.551	0.628
2	0.711	0.768	0.738
3	0.852	0.742	0.793
4	0.735	0.942	0.826
5	0.871	0.839	0.855
6	0.759	0.783	0.771
7	0.679	0.918	0.781
8	1.000	0.859	0.924
9	0.852	0.757	0.801
10	0.719	0.939	0.814
Average	0.791	0.808	0.792

Based on the Table-1, the average classification performance reached 0.791 in accuracy, 0.808 in precision and 0.792 in f-measure. The highest accuracy was achieved by the Subject 8, with perfect classification rate. The precision and f-measure were recorded at 0.859 and 0.924 respectively. On the other hand, the lowest accuracy was yielded by the Subject 7. It is recorded at 0.679 in accuracy, 0.918 in precision and 0.781 in f-measure.

Furthermore, the classification performance is further tested with different window size threshold. The step size of training pool is linear increased by 10%, from 20% up to 90%. The classification performance with different window size threshold is shown in Figure-1.

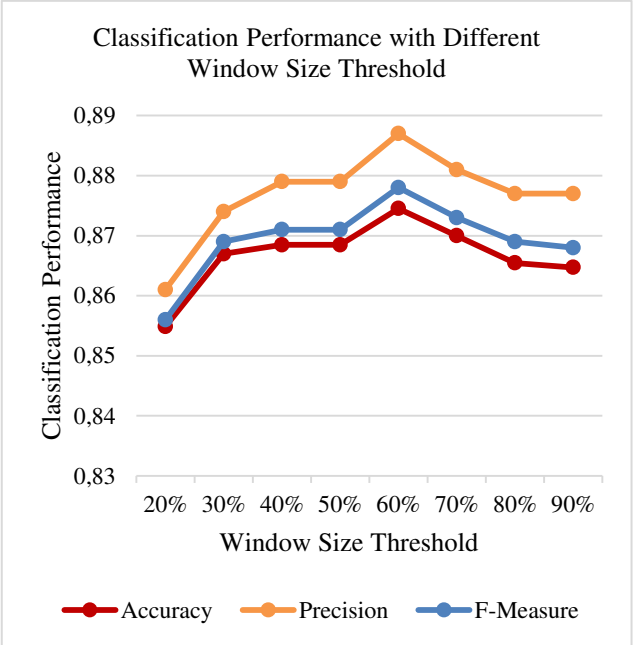

Figure-1. Classification performance with different window size threshold.

Based on the figure above, the accuracy with 20% window size threshold had increased 0.136, 0.053 and 0.064 for the accuracy, precision and f-measure respectively. The classification performance was increased drastically when the window size threshold was increased to 30%.The classification performance recorded at 0.867 in accuracy, 0.874 in precision and 0.869 in f-measure. After that, the classification performance showed slightly changes from the window size threshold of 30% to 50%.The optimum classification performance was showed in the window size threshold of 60%. It is recorded at 0.875 in accuracy, 0.887 in precision and 0.878 in f-measure. However, the classification performance started to decrease after this. The classification performance with the window size threshold of 70% decreased to 0.870 in accuracy, 0881 in precision and 0.873 in f-measure. The classification performance decreased continuously for the next 10% of window size threshold, recorded at 0.865 in accuracy, 0.877 in precision and 0.869 in f-measure. With the 90% of window size threshold, the classification performance showed slightly decreased in accuracy and f-measure while precision remained the same. This is proven that the greater number of training objects do not help to improve the classification performance.

From this experiment, we can observe that the incremental learning plays important role in brainprint identification. The incremental learning includes the new EEG signals characteristics for the brainprint identification. The classification performance was improving when the size of training pool is getting larger. However, the classification performance was decreased after reached an optimum threshold. It might because huge number of the training objects result in more uncertainty in the knowledge base. The new objects included in the training pool might not representative and misleading. Hence, the classification performance was degraded.

5. CONCLUSIONS

In this paper, we have investigated the classification performance using different window size threshold for brainprint identification. The incremental learning plays crucial role to capture the new information and changes from the EEG signals due to the non-stationarity property. However, the experiment had proven that the FIFO incremental update strategy is less promising. It is not a good incremental update strategy for the training pool to include all the new objects whenever it is available. It is because not all the new objects are distinct and representative. The greater number of training objects increased the uncertainty level in the training pool. The increased of uncertainty in the training pool will jeopardized the classification performance. Thus, the future work should focus on the selection of new objects to be included to the training pool instead of included all the new objects.

ACKNOWLEDGEMENT

The authors would like to express their appreciation to the Universiti Teknikal Malaysia Melaka (UTeM) for providing the UTeM Zamalah Scheme scholarship. Besides, the authors would also like to thank UTeM for the research facilities support.

REFERENCES

[1] M. Del Pozo-Banos, J. B. Alonso, J. R. Ticay-Rivas, and C. M. Travieso. 2014. Electroencephalogram Subject Identification: A Review. Expert Syst. Appl. 41(15): 6537-6554.

[2] B. C. Armstrong, M. V. Ruiz-Blondet, N. Khalifian, K. J. Kurtz, Z. Jin and S. Laszlo. 2015. Brainprint: Assessing the uniqueness, collectability and permanence of a novel method for ERP biometrics. Neurocomputing. 166: 59-67.

[3] S. Yang. 2015. The Use of EEG Signals for Biometric Person Recognition. University of Kent.

[4] M. V. Ruiz-Blondet, Z. Jin and S. Laszlo. 2017. Permanence of the CEREBRE brain biometric protocol. Pattern Recognit. Lett. 95: 1339-1351.

[5] Z. Y. Ong and Z. Ibrahim. 2018. Power Spectral Density Analysis for Human EEG- based Biometric Identification. in 2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA). pp. 1-6.

[6] Q. Gui, M. V. Ruiz-blondet and S. Laszlo. 2019. A Survey on Brain Biometrics. ACM Comput. Surv. 51(6): 112: 1-38.

[7] J. Ortega, K. Martín-Chinea, J. F. Gómez-González and E. Pereda. 2020. Biometric Person Authentication Using a Wireless EEG Device. in International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning. pp. 615-620.

[8] S.-H. Liew, Y.-H. Choo, Y. F. Low and S. H. Chong. 2020. Investigation of Alpha and Beta Band for Brainprint Authentication with Auditory Distractor. Int. J. Comput. Inf. Syst. Ind. Manag. Appl. 12: 14-22.

[9] B. Singh, S. Mishra and U. S. Tiwary. 2015. EEG Based Biometric Identification with Reduced Number of Channels. in 2015 17th International Conference on Advanced Communication Technology (ICACT). pp. 687-691.

[10] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. Mohamad Saad and A. S. Malik. 2017. Classification of EEG Signals Based on Pattern Recognition Approach. Front. Comput. Neurosci. 11: 1-12.

[11] P. Joshi and P. Kulkarni. 2012. Incremental Learning: Areas and Methods - A Survey. Int. J. Data Min. Knowl. Manag. Process. 2(5): 43-51.

[12] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, A. Rakotomamonjy and F. Yger. 2018. A Review of Classification Algorithms for EEG-based Brain-Computer Interfaces : A 10-year Update. J. Neural Eng. 15(3): 1-55.

[13] S. Yang and F. Deravi. 2017. On the Usability of Electroencephalographic Signals for Biometric Recognition: A Survey. IEEE Trans. Human-Machine Syst. 47(6): 958-969.

[14] S. H. Liew, Y. H. Choo, Z. I. Mohd Yusoh and Y. F. Low. 2016. Incrementing FRNN Model with Simple Heuristic Update for Brainwaves Person Authentication. in IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). pp. 115-120.

[15] S. H. Liew, Y. H. Choo, Y. F. Low and Z. I. Mohd Yusoh. 2018. EEG-based biometric authentication modelling using incremental fuzzy-rough nearest neighbour technique. IET Biometrics. 7(2): 145-152.

[16] G. Safont, A. Salazar, A. Soriano and L. Vergara. 2012. Combination of multiple detectors for EEG based biometric identification/authentication. in 2012 IEEE International Carnahan Conference on Security Technology (ICCST). pp. 230-236.

[17] J. F. D. Saa and M. S. Gutierrez. 2010. EEG Signal Classification Using Power Spectral Features and linear Discriminant Analysis : A Brain Computer Interface Application. in 8th Latin American and Caribbean Conference for Engineering and Technology. pp. 1-7.

[18] Y. F. Low and D. J. Strauss. 2011. A performance study of the wavelet-phase stability (WPS) in auditory selective attention. Brain Res. Bull. 86: 110-117.

[19] M. A. Hall. 2000. Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning. in Proceeding ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning. pp. 359-366.