Automatic feature extraction from wearable sensor data by use of machine learnings
Keita SATO, Masafumi CHIDA, Yoshihiro HAYAKAWA and Nahomi M. FUJIKI
National Institute of Technology, Sendai College Hirose Campus
4-16-1 Ayashi-chuo, Sendai, 989-3128, JAPAN
*Corresponding Author: fujiki@sendai-nct.ac.jp

Abstract
Feasibility of machine leaning techniques for use of automatic detection and analysis of human behavior has been discussed. We focused on the analysis of swing motions in ping-pong rallies based on the acceleration sensor data of a Smartwatch and to identify slight difference contained in the motion that is unnoticeable by seeing. It has been discussed the feature extraction ability of PCA and auto encoder based on a multilayered neural network combined with clustering methods and shown the possibility of automatic detection of small scale difference in swing motions.

Keywords: Sensor devices, Principal Component Analysis, Multilayered Neural Network.

1. Introduction
Great development of data processing technology and data processing ability of computers enable us to collect big data easily, analyze them, and extract meanings concealed inside of such big data. On the other hand, recent miniaturization technology of sensor devices realizes wearable devices which can collect many kinds of biodata and physical data from human activities anytime and/or anywhere. If we can monitor and analyze human activities on real time without giving the person any stress, it becomes feasible to offer various kinds of services related to human activities. Furthermore, machine learning techniques including neural networks or deep learning systems provides us a way of analyzing such collected sensor data based on human activities in microscopic scale that we can’t identify visually.

In this paper, we focus on an auto-analysis system of a sport activity that is consisted of repetitive motions such as swings in ping-pong rallies. We expect that use of machine learning techniques makes automatic detecting of any minimal change in such swing motions possible. A smartwatch attached on the wrist of a player is able to collect sensor data from a player during his/her play. This smartwatch is equipped various functions such as acceleration sensor, linear acceleration sensor and gyro sensor. We collected the acceleration data of a player’s wrist during ping-pong rallies and applied various machine learning techniques to analyze them. At the same time, the video image of the player is recorded by a camera, so that we can verify certain meanings of differences in swing motions extracted from the sensor data.

We adopted Principal Component Analysis (PCA), Neural Network auto-encoding (NN) and clustering methods. Some of machine learning techniques can map multi-dimensional data onto low dimensional feature space and help us to identify minimal characteristic concealed in such data.

We use a Smartwatch M600 (POLAR co.) that can collect acceleration sensor data along three axes of forehand swings of a player who wears the watch on his right wrist. The sampling frequency of the smartwatch is 100Hz. At the same time, the foreside and right-hand side movements of the player are recorded by video cameras (HDR-AS300 Sony co. and GZ-E765-R JVC co.). Since one full swing takes roughly 1280msec, we provide a data point in 384 (128 acceleration data for one swing times 3 axes) dimensions corresponding to one full swing. Altogether, a set of 633 data points is used to analyze forehand ping-pong swing motions.

The directions of x, y and z axes defined on the smartwatch are indicted in Figure 1 and acceleration sensor raw data along these three different axes are shown in Figure 2. It is obvious that the amplitude of acceleration
along x axis or y axis in Fig.2 is larger and periodic while the change in the acceleration data along z axis is rather small.

![Fig.1. Smartwatch and the definition of 3 axes](image1)

![Fig.2. Acceleration raw data along 3 axes](image2)

![Fig.3. Similar acceleration data in different phase](image3)

2. PCA analysis for acceleration sensor

Principal Component Analysis treats multivariate dataset in order to extract the largest possible variance with its most informative viewpoint. We applied PCA to the acceleration sensor dataset and analyze the meaning of each major component.

Actual period length for each swing may change slightly one after another as shown in Figure 3. Therefore cutting out sensor data in a constant time period yields phase shift influence into the analysis result. In order to eliminate such phase effect, we introduced FFT (Fast Fourier Transfer) before applying PCA to the dataset.

The 1st PCA components (PC1) and the 2nd components (PC2) from the result of FFT+PCA are plotted on the figure 4 (a). Then, we apply X means method to these PCA results. X means method classified the data set into 4 clusters as shown in Fig. 4 (a). As comparison, the PCA result without FFT pre-treatment is shown in Figure 4 (b) too. Data points in one of 4 clusters in Fig. 4 (a) are circularly wide-spread in the figure 4 (b). Therefore, it can be understood that FFT treatment has eliminated phase effects in the dataset. Typical feature of acceleration raw data corresponding to each cluster is shown in the figure 5. The 4 clusters get into line along PC1 axis and data of each cluster spreads into PC2 direction. Let’s call the cluster of lowest PC1 value Cluster 1 and then Cluster 2 for the next value and so on.

![Fig.4. Results of PCA+FFT (a) and PCA only (b). PCA+FFT results classified in 4 clusters.](image4)
Lower the PC1 value gives higher the amplitudes of x and y directional accelerations.

Compare to the feature that x and y directional changes of acceleration are quite large amount, the feature of Cluster 4 shows very small change of acceleration data in any direction. The video images are examined to find out a meaning of spread in PC2 component.

We could recognize that when PC2 value is large the player hit a ball at slightly far from the camera that means the player hit a ball near center of the table and when PC2 value is smaller the player is closer to right corner of the table. However we couldn’t find any obvious correlation between sensor data and PC2 values.

A layered neural network is well-known for its excellent learning ability. We also tried to classify a set of sensor data using an auto-encoder consisted of neural network layers. We can realize identity mapping on a multilayered neural network if we make the network supervised learning by using the same dataset as an input and training data. In this way, the information of a 384-dimensional sensor dataset can be compressed into a low dimensional feature space created in the middle layer of a multilayered NN. The basic structure of an auto-encoder NN used in this research is shown in Figure 6.

Datasets of 300 swings out of 633 original swings was used for supervised learning. Numerical experiment with three different initial states of the NN was made for 30,000 learning steps. We adopted Momentum SGD (Stochastic Gradient Decent) with the leaning constant parameter 0.5 and the momentum parameter 0.8 as the supervised learning algorithm. Each 384-dimension sensor data was reduced into two dimensional feature space as shown in Figure 5. Applying X means method to these reduced

Fig. 5 Sample acceleration raw data of x,y and z directions corresponding to the clusters in Fig. 4 (a) obtained by X-means

3. Auto-encoder multilayered neural network for acceleration sensor data
dimensional data resulted three clusters in the feature space. In order to figure out the meaning of each cluster we tried to map out the data points of each cluster to the data points on another feature space obtained from supervised learning with different initial condition. Shapes of clusters on each feature space are different but almost all data points belonged to each cluster are placed together in the other feature space as well. However we couldn’t find any direct relation between clusters of PCA space and NN feature space.

Close examination of sensor data and video images corresponding to each cluster on the feature space didn’t lead to any firm conclusion that could explain any specific meaning of clusters. Because of the nonlinearity of neural network data processing, even if it is possible to extract certain characters submerged in the sensor data but it seems difficult to identify such difference visually.

In order to observe the phase shift effect of raw data on the result in NN feature space, we tried to map some points from one cluster in the feature space onto PCA space. Because these points are spread rather circularly in the PCA space, it may be no phase effect included in the clustering result of NN feature space.

4. Analysis of correlation between ball placements and sensor data

Instead of concentrating on a player’s movement in video images, we proceed with further analysis of the feature extracted data based on the placement of a ping-pong ball before and after a swing. Focusing on a ball position extracted from video images may help us to find out the meaning of extracted features in PCA and NN feature spaces. We divided a table of player’s own side court into 4 areas and 9 areas for an opponent side court as shown in Figure 8. The player in this experiment is right handed and we collected acceleration sensor data of 290 forehand swings continuously. We chose linear acceleration data of a wearable device taken for one second just after a ball landed on the player’s own side court before a swing and defined it as a set of data corresponding to one swing. In this way, the phase effect depending on the choice of a starting point of each swing and gravitational force effect should be minimized. The wearable device, Smartwatch M600, can measure linear accelerations of swing motions with each 20msec interval. We collected 44 data points of sampled linear acceleration data just after the ball landed on the table of player’s own side and defined it as one swing dataset and consequently provided a 132 (44 × 3 directions) dimensional dataset representing one swing motion. In this way, we

Fig.7. (a) The 3 clusters of data in the feature space classified by X-means. (b), (c) Results of mapping the clustered points onto two other feature spaces of NNs that started its learning from different initial states.
provided 290 swing datasets.

Feature extraction of swing motions is tried by using Principal Component Analysis (PCA) and an autoencoder using layered neural network (NN). PCA and NN reduced the swing dataset into two-dimensional feature space shown an example in Figure 9 (a) and (b). The k-means method classifies data points in the 2-dimensional feature space into k clusters.

We examine ball placements for each cluster and estimate the probability of a course for a return ball. Some of results are shown in Figure 10. The average probability over clusters shown in the figure indicates the tendency that most of returns were hit from the area I (132 out of 290) or II (86 out of 290) and landed in the area E, F or I. However, some of clusters show characteristic feature as shown in Figure 10.

Expectation value of area prediction for ball placements is defined as the function of the number of clusters by the following formula.

\[ E_k = \sum_{x=1}^{IV} \sum_{c=1}^{k} P(x, c) \max_i T_{xc}(i) \]

Here, \( k \) is the number of clusters, \( P(x, c) \) is the joint probability that a ball corresponding to sensor data belonged to a cluster \( c \) lands on an area \( x \) among four different areas from I to IV. \( T_{xc}(i) \) is the conditional probability that the ball in the cluster \( c \) is returned from an area \( x \) in the player’s side court and then lands at an area \( i \).
among 9 divided areas in an opponent side table.

Expectation values estimated for results of PCA and an auto-encoder NN are shown in Figure 11. Naturally better prediction rate is obtained increasing the number of k. Data compressed into 2-dimensional space by NN indicated slightly higher expectation value for each number of clusters compared to PCA results.

4. Conclusions

Feasibility of machine learning techniques for use of automatic detection and analysis of human behavior has been discussed in this paper. In order to testify the ability of various machine learnings, we focused on the analysis of swing motions in ping-pong play that limits its action and area of movement so that it seems relatively easy to analyze player’s behavior. We tried to classify acceleration data collected from swing motions and identify slight difference contained in the motion that is unnoticeable by seeing. Wearable devices allow us to measure acceleration changes in swing motions on time and machine learnings make us possible to analyze deep implication in the acceleration data.

At same time, the video image of player’s swing motions has been recorded and used to support understanding the meaning of result obtained by machine learning analysis.

There are two ways of machine learnings under consideration in this paper. One is PCA and the other is an auto encoder based on multilayered neural network.

PCA applied to the data after reducing the phase effect by FFT indicated the 1st component depends on the difference in the amplitude of acceleration indicating the level of swing speed, and the 2nd component depends on the body movement of a player.

On the other hand, the auto-encoder of multilayered neural networks could reduce the acceleration data into two-dimensional feature space. X-means clustering method classified 633 swings in the feature space into three clusters. Numerical experiments on the auto encoder have been done for three different initial states and we have three different clustering patterns corresponding to different initial states.

Mapping each point in one feature space onto other feature space obtained by another learning process with a different initial condition suggested the existence of one to one correspondence between clusters in different feature spaces obtained by the auto encoder NN.

It was difficult to find clear relationship between points in a same cluster but the neural network has successfully extracted certain characteristic difference
among those swing datasets and found out the similarity of some kind.

Instead of focusing on correlation between video image and acceleration data of swing motions, analysis focusing on ping-pong ball placements was discussed. Dependency of each cluster on ball placements is examined with changing a cluster number k and the expectation values for predicting a course of a ball are estimated. The result obtained by an auto-encoder showed slightly better expectation values than the result of PCA.

It is shown that machine learning techniques are possible to specify human behavior in small scale automatically. However, it is required further investigation based on different points of view to establish a method that can fully identify the meaning of such small difference detected by the machine learning.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant Numbers 17K00356 and 16K01300.

References