International Journal of Innovative Computing, Information and Control
Volume **15**, Number **1**, February **2019**

pp. **199–208**

A NOVEL DEEP LEARNING OPTIMIZATION ALGORITHM FOR HUMAN MOTIONS ANOMALY DETECTION

Yuto Omae[1], Masaya Mori[2], Takuma Akiduki[3] and Hirotaka Takahashi[2]

[1]Department of Electrical Engineering
National Institute of Technology, Tokyo College
1220-2, Kunugida, Hachioji, Tokyo 193-0942, Japan
yuto.omae@gmail.com

[2]Department of Information and Management Systems Engineering
Nagaoka University of Technology
1603-1, Kamitomioka, Nagaoka, Niigata 940-2188, Japan
hirotaka@kjs.nagaokaut.ac.jp

[3]Department of Mechanical Engineering
Toyohashi University of Technology
1-1, Hibarigaoka, Tenpakucho, Toyohashi, Aichi 441-8580, Japan

Received May 2018; revised September 2018

Abstract. *Recently, there are many researches to detect the anomaly of human motions by using the machine learning and inertial sensors. In general, the individual differences exist in human motion by the height, body weight, habits and so on. Classification models based on the deep learning have often high quality. However, the general deep learning optimization algorithms do not consider the individual differences in human motions. By the reason, classification model based on the algorithms does not guarantee to take into account the individual differences. Therefore, we propose a novel deep learning optimization algorithm for human motions' anomaly detection from the data of the inertial sensor. The reliability of the proposed algorithm is also confirmed by the collected dataset.*
Keywords: Inertial sensor, Human motion, Deep learning, Mathematical optimization, CHI-FS evaluation function

1. **Introduction.** By using the machine learning and inertial sensors, many researches of human activity recognition exist. For example, Morris et al. [1] developed a muscle training motion recognition system by using a linear support vector machine. Bao and Intille [2] developed an ordinal activities recognition (e.g., walking, sitting, and running) by using C4.5 decision tree. Omae et al. [3] developed a swimming style classification model (butterfly, front crawl, backstroke and breaststroke) by using random forest classifier. Jensen et al. [4] and Kobayashi et al. [5] developed a swimming motion (stroke and turn motion) recognition model. Moreover, there are the human activity recognition models based on the deep learning from the data of the inertial sensors (e.g., Alsheikh et al. [6], Ordonez and Roggen [7] and Ravi et al. [8]). In general, human motions include the individual differences among the humans, such as the height, body weight, and habits. Thus, the signals from the inertial sensors include these differences.

Omae and Takahashi [9] pointed out that we should consider the individual differences if we use the machine learning algorithm. However, general feature selection algorithms (e.g., out-of-bug error method [10], reliefF [11] and minimum reference set [12]) to develop high quality classifier do not consider them. Thus, Omae and Takahashi proposed a

DOI: 10.24507/ijicic.15.01.199

feature selection algorithm considering them for the classification problems [9]. Moreover, Mori et al. extended their algorithm [9] for the anomaly detection problems [13, 14, 15]. As their results, the classification quality had higher score than that using other feature selection algorithms. The algorithm was developed to search for the effective features from many features of the time and/or frequency domain of inertial sensors. It cannot use the optimization of deep learning-based classification or anomaly detection model.

Therefore, in this paper, we propose a deep learning optimization algorithm considering the individual differences by extending previous algorithm. Our target problem is the human motions' anomaly detection and target model is the Convolutional Neural Network (CNN) as the deep learning.

The paper is organized as follows. In Section 2, we briefly explain the CNN-based feature extraction from the inertial sensors and its model parameters. In Section 3, we explain our proposed method of the CNN learning algorithm considering the individual differences of human motions. In Section 4, we show the results of the experiment to confirm the reliability of the proposed algorithm. Section 5 is devoted to a summary.

2. **CNN-Based Feature Extraction and Its Model Parameters.** Figure 1 shows the schematic view of the CNN-based feature extraction from the inertial sensors and its model parameters.

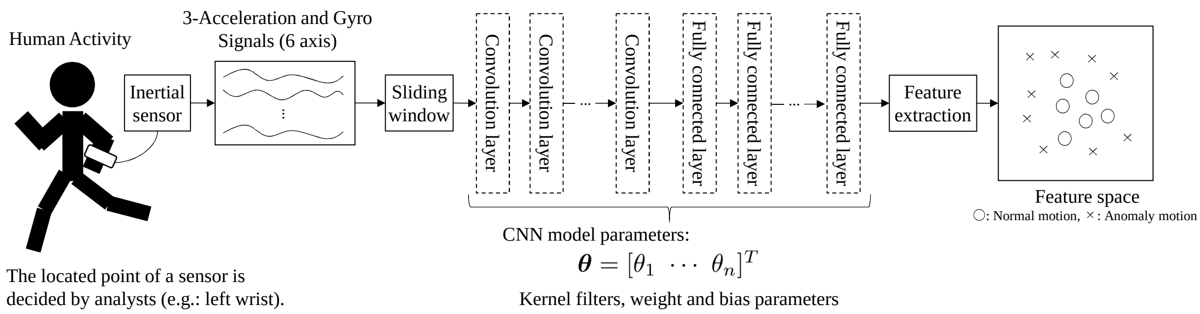

FIGURE 1. CNN-based feature extraction and model parameters $\boldsymbol{\theta}$

Initially, we attach the sensors to a subject's body. The located points are the back waist, wrist, and so on (e.g., the located point of the subject in Figure 1 is the left wrist). The inertial sensors measure the signals (X, Y, Z-axes acceleration and gyro data) from human motions. The signals processed by the sliding window [16] are inputed into the CNN. They are transformed into the low dimensional matrix based on the kernel filters. After that, the feature extraction is performed via the fully connected layers. The CNN has many model parameters to carry out the feature extraction (e.g., kernel filter, weight and bias parameters). We line up model parameters θ_α and make the column vector that is expressed by,

$$\boldsymbol{\theta} = [\theta_1 \ \cdots \ \theta_n]^T, \tag{1}$$

where n is the number of model parameters. Then, the feature vector based on the CNN is expressed by,

$$\boldsymbol{f}(\boldsymbol{\theta}) = [f_1(\boldsymbol{\theta}) \ \cdots \ f_m(\boldsymbol{\theta})]^T, \tag{2}$$

where m is the number of output neurons of the CNN. By definition of Equation (2), the feature vector $\boldsymbol{f}(\boldsymbol{\theta})$ depends on model parameter $\boldsymbol{\theta}$.

3. **Optimization Algorithm.** In this section, we describe our proposed method of the CNN learning algorithm considering the individual differences of human motions. The algorithm aims to detect the anomaly motion, which is independent of the individual differences, from many subjects. In the proposed algorithm, we use a Variance and Correlation (VC) evaluation function and a Consideration of Human motion's Individual differences-based Feature Space (CHI-FS) evaluation function [13, 14, 15]. By using the VC and CHI-FS evaluation functions, the CNN model parameters $\boldsymbol{\theta}$ are optimized based on the Gradient Descent (GD).

3.1. **Variance and correlation evaluation function.** We explain the VC evaluation function. It evaluates the variances of the features and correlations among the features of i-th subject in the feature space based on the CNN. The schematic view of the VC evaluation function is shown in Figure 2(a). Initially, we define,

$$o^i_{\delta_1}(\boldsymbol{\theta}) = \sum_{j\in\{1,\ldots,m\}} \frac{\mathrm{Var}\left(f^i_{j,\mathrm{nor}}(\boldsymbol{\theta})\right)}{\mathrm{Var}\left(f^i_{j,\mathrm{ano}}(\boldsymbol{\theta})\right)} + \frac{1}{2m^2-m} \sum_{j\in\{1,\ldots,m\}} \sum_{k\in\{1,\ldots,m\}} \sum_{l\in\{\mathrm{nor},\mathrm{ano}\}} \left|\mathrm{Cor}\left(f^i_{j,l}(\boldsymbol{\theta}), f^i_{k,l}(\boldsymbol{\theta})\right)\right| - m, \tag{3}$$

where $\mathrm{Var}\left(f^i_{j,\{\mathrm{nor},\mathrm{ano}\}}(\boldsymbol{\theta})\right)$ means the variance of i-th subject's feature $f^i_j(\boldsymbol{\theta})$ of class label $\{\mathrm{nor}, \mathrm{ano}\}$. "nor" and "ano" show the normal and anomaly motions, respectively (e.g., "nor": walking, "ano": falling down, stumbling and so on). Moreover, $\mathrm{Cor}(f^i_{j,l}(\boldsymbol{\theta}), f^i_{k,l}(\boldsymbol{\theta}))$ means the correlation of i-th subject's feature $f^i_j(\boldsymbol{\theta})$ and $f^i_k(\boldsymbol{\theta})$ of class label $l \in \{\mathrm{nor}, \mathrm{ano}\}$. The range of $o^i_{\delta_1}(\boldsymbol{\theta})$ is $[0, \infty)$. Equation (3) represents the evaluation value of i-th subject's feature space. The first term of Equation (3) consists of the variance of the normal and anomaly data in the feature space. If the variance of normal data is small and the variance of anomaly data is high, the first term achieves small value. The second term

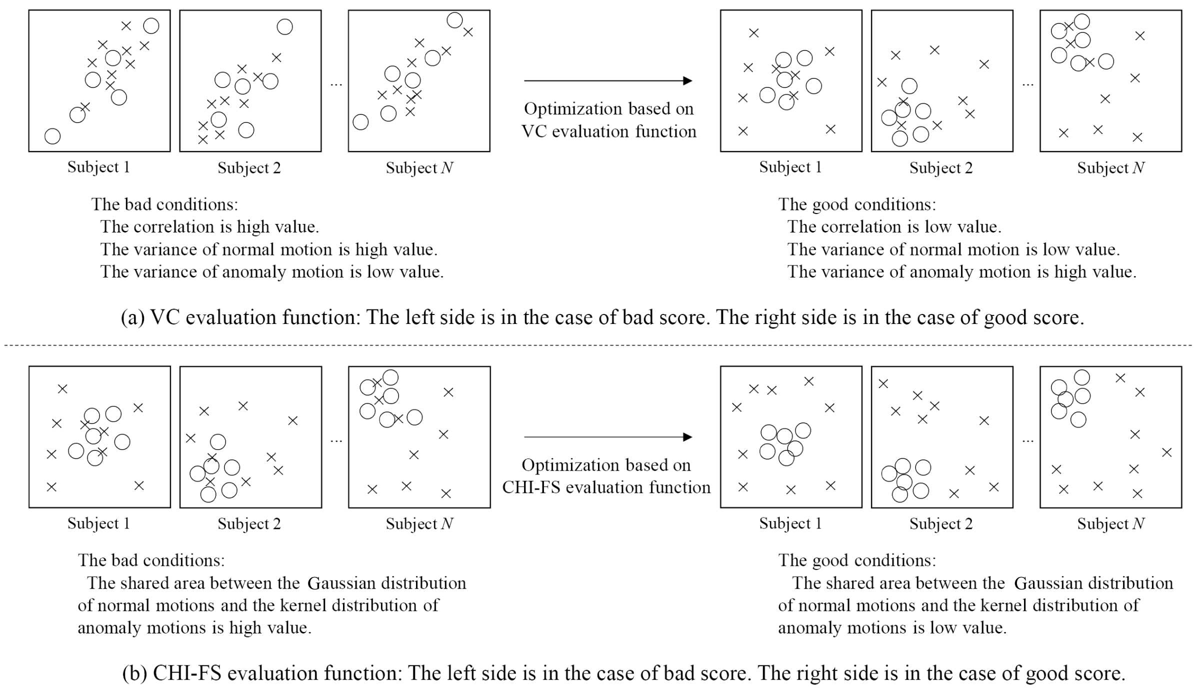

Figure 2. The relationship between the evaluation functions and the feature space

of Equation (3) consists of the correlation of the normal and anomaly data in respective features value. Note that the number of the correlation in the second term except for the correlation between the same features is $2m^2 - m$, then the reason of division $2m^2 - m$ in the second term is the normalization. It is desirable that we do not have the correlation relationship between the respective features for the feature space. If the correlation is small, the second term is small. The third term has an effect to remove the correlation between the same feature. In the case of the correlation between the same feature, this value is one. Since the number of their cases in the second term is m, we subtract m in the third term. Therefore, if $o^i_{\delta_1}(\boldsymbol{\theta})$ has small value, it means a good feature space. In contrast, if $o^i_{\delta_1}(\boldsymbol{\theta})$ has high value, it means a bad feature space.

From the calculation of Equation (3), the evaluation values of the respective subject can be obtained. After that, we transform them into the evaluation value of all subjects by,

$$O_{\delta_1}(\boldsymbol{\theta}) = w_1 \mathrm{Mean}(o_{\delta_1}(\boldsymbol{\theta})) + (1 - w_1)\mathrm{Std}(o_{\delta_1}(\boldsymbol{\theta})), \quad w_1 \in [0, 1], \tag{4}$$

where $\mathrm{Mean}(o_{\delta_1}(\boldsymbol{\theta}))$ and $\mathrm{Std}(o_{\delta_1}(\boldsymbol{\theta}))$ are all subjects' mean and standard deviation of $o^i_{\delta_1}(\boldsymbol{\theta})$, $\forall i$, respectively. w_1 is a weight parameter that is range from zero to one. If the value of w_1 is higher than 0.50, it treats the mean value of all subjects as important. In contrast, if w_1 is smaller than 0.50, it treats the standard deviation of all subjects as important. In other words, the role of w_1 is an adjustment parameter between the mean value and standard deviation. The evaluation value $O_{\delta_1}(\boldsymbol{\theta})$ achieves good value if the variances of the normal data of all subjects are low, the variances of anomaly data of all subjects are high and the correlations of the respective features of all subjects are low.

3.2. **CHI-FS evaluation function.** The CHI-FS evaluation function is developed by Mori et al. [13, 14, 15] and evaluates the integrated error risk between the normal and anomaly data in the feature space. The schematic view of the CHI-FS evaluation function is shown in Figure 2(b). The CHI-FS evaluation function considers the individual differences of the respective subject and the output value is good if the error risk of many subjects achieves small error risk. In other words, if the error risk of only some subjects achieves small, the evaluation value is not small. In this case, the output of the CHI-FS evaluation function should be bad value. Therefore, we can search for a good feature space for many subjects by using this function. For example, Mori et al. [14, 15] searched for the best feature space for the aimless detection from the acceleration and gyro data of inertial sensors attached on the right and left wrists. However, in this case study, they did not use deep learning. In other words, the CHI-FS evaluation function in previous research was used to perform the feature selection.

In this paper, we extend this CHI-FS evaluation function to use in deep learning. We calculate the class overlapping function expressed by,

$$D^i(\boldsymbol{f}(\boldsymbol{\theta})) = \prod_{j \in \{\mathrm{nor}, \mathrm{ano}\}} P^i_j(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta})), \tag{5}$$

where,

$$\begin{aligned} &P^i_{\mathrm{nor}}(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta})) \\ &= \frac{1}{(2\pi)^{m/2}\sqrt{|\mathrm{Cov}(\boldsymbol{f}^i(\boldsymbol{\theta}))|}} \\ &\quad \times \exp\left\{-\frac{1}{2}\left(\boldsymbol{x} - \mathrm{Mean}\left(\boldsymbol{f}^i(\boldsymbol{\theta})\right)\right)^T \mathrm{Cov}\left(\boldsymbol{f}^i(\boldsymbol{\theta})\right)^{-1}\left(\boldsymbol{x} - \mathrm{Mean}\left(\boldsymbol{f}^i(\boldsymbol{\theta})\right)\right)\right\}. \end{aligned} \tag{6}$$

$\mathrm{Cov}(\boldsymbol{f}^i(\boldsymbol{\theta}))$ is the $m \times m$ size variance-covariance matrix of i-th subject's feature vector $\boldsymbol{f}^i(\boldsymbol{\theta})$. Mean $\left(\boldsymbol{f}^i(\boldsymbol{\theta})\right)$ is the m dimensional mean vector of i-th subject's feature vector $\boldsymbol{f}^i(\boldsymbol{\theta})$. And,

$$P^i_{\mathrm{ano}}(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta})) = \frac{1}{N^i}\sum_{l=1}^{N^i} \prod_{z\in\{1,\dots,m\}} \frac{1}{h^i_z(\boldsymbol{\theta})} K\left(\frac{x_z - f^{i,l}_z(\boldsymbol{\theta})}{h^i_z(\boldsymbol{\theta})}\right),\ K(a) = \frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}a^2\right\}, \quad (7)$$

where $f^{i,l}_z(\boldsymbol{\theta})$ is the l-th real obtained data of i-th subject's feature $f^i_z(\boldsymbol{\theta})$, N^i is the number of i-th subject's samples of anomaly data and $K(a)$ is the kernel function, and we use Gaussian kernel. $h^i_z(\boldsymbol{\theta})$ is the i-th subject's bandwidth [17] of z-th feature $f^i_z(\boldsymbol{\theta})$ and expressed by,

$$h^i_z(\boldsymbol{\theta}) = 1.06\ \mathrm{Std}\left(f^i_z(\boldsymbol{\theta})\right)(N^i)^{-1/5}, \quad (8)$$

where $\mathrm{Std}\left(f^i_z(\boldsymbol{\theta})\right)$ is the i-th subject's standard deviation of feature $f^i_z(\boldsymbol{\theta})$. This bandwidth $h^i_z(\boldsymbol{\theta})$ is defined by Sheather [18].

$D^i(\boldsymbol{f}(\boldsymbol{\theta}))$ of Equation (5) means the overlapping amount on the feature space by the features $\boldsymbol{f}(\boldsymbol{\theta})$ of i-th subject. $P^i_{\mathrm{nor}}(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta}))$ of Equation (6) means the probability distribution of the data labeled by "nor" of i-th subject. We assume the Gaussian distribution as $P^i_{\mathrm{nor}}(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta}))$ because of the normal human motion (e.g., walking). $P^i_{\mathrm{ano}}(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta}))$ of Equation (7) means the probability distribution of the data labeled by "ano" of i-th subject. We assume the non-parametric distribution as $P^i_{\mathrm{ano}}(\boldsymbol{x}; \boldsymbol{f}(\boldsymbol{\theta}))$ because of the anomaly human motion (e.g., falling down, stumbling and so on).

After that, we calculate the error risk of i-th subject expressed by,

$$o^i_{\delta_2}(\boldsymbol{\theta}) = \int\cdots\int_L D^i\left(\boldsymbol{f}(\boldsymbol{\theta})\right) d\boldsymbol{f}(\boldsymbol{\theta}), \quad (9)$$

where L is an integration range and we set $L = [0, 1]$. It is required the standardization of the feature vector's range from zero to one. As the result, we can get the risk of the miss-classification on the features $\boldsymbol{f}(\boldsymbol{\theta})$ of i-th subject.

After we obtain $o^i_{\delta_2}(\boldsymbol{\theta})$ of all subjects, we calculate the evaluation value defined by,

$$O_{\delta_2}(\boldsymbol{\theta}) = w_2\mathrm{Mean}(o_{\delta_2}(\boldsymbol{\theta})) + (1 - w_2)\mathrm{Std}(o_{\delta_2}(\boldsymbol{\theta})), \quad w_2 \in [0, 1], \quad (10)$$

where $\mathrm{Mean}(o_{\delta_2}(\boldsymbol{\theta}))$ and $\mathrm{Std}(o_{\delta_2}(\boldsymbol{\theta}))$ are the all subjects' mean and standard deviation of $o^i_{\delta_2}(\boldsymbol{\theta})$, $\forall i$, respectively. w_2 is a weight parameter that is range from zero to one. Note that Equation (10) has the same meaning of Equation (4).

3.3. **Optimization.** We carry out the optimization of the CNN model parameters based on the VC and CHI-FS evaluation functions. If the value of evaluation functions achieves low score, the CNN performs the good feature extraction such as the right side figures of Figure 2. Therefore, as the best model parameters $\boldsymbol{\theta}^{\mathrm{opt}}$, we defined,

$$\boldsymbol{\theta}^{\mathrm{opt}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}}\, O_{\delta_c}(\boldsymbol{\theta}), \quad c \in \{1, 2\}. \quad (11)$$

However, there are many model parameters (i.e., n is very large value). Because we cannot find a minimum value of it, we adopt the suboptimal solution as the CNN model parameters by using the GD method. In other words, we update the CNN model parameters by using,

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \gamma\nabla O_{\delta_c}(\boldsymbol{\theta}), \quad \nabla = \left[\frac{\partial}{\partial\theta_1}\cdots\frac{\partial}{\partial\theta_n}\right]^T, \quad (12)$$

where $\boldsymbol{\theta}^t$ means the t-th parameters of $\boldsymbol{\theta}$, and γ means the learning coefficient.

Initially, we use the VC evaluation function ($c = 1$) to obtain the CNN model parameters for the feature extraction such as the right side figures of Figure 2(a). After the converged VC evaluation function, we use the CHI-FS evaluation function ($c = 2$). By switching the evaluation functions, the CNN aims to extract the features such as the right side figures of Figure 2(b).

4. **Experiment.**

4.1. **Outline.**

We perform an experiment to confirm the reliability of the proposed algorithm. We use the data form the swimming situation. The number of subjects is six who are the university students. A single inertial sensor is attached on the back waist. We used an inertial sensor made by Sports Sensing Co., Ltd with the following specifications: accelerations (± 5 G); angular velocities (± 1500 dps); sampling frequency (100 Hz); weight 20 g; size $67 \times 26 \times 8$ mm [19]. After starring measurement of inertial sensor data, the subject performs walking and jumping and so on (various human motions). After that, they carry out the butterfly as swimming with the full force (distance: 50 m). Finally, they leave from the pool and perform walking, jumping, and so on (various human motions).

We took the movie of their motions by using video camera (30 fps). Then, we gave the label "nor" to butterfly motions and label "ano" to those except butterfly motions. In other words, the normal motion means the butterfly and the anomaly motions mean the various motions (walking, jumping, leaving motion from the pool and so on).

4.2. **Result and discussions.**

The six subjects were divided into two groups. One of them is the training data for the CNN and the other is the test data. The training data is the dataset for the parameter optimization and consists of three persons' data. The test data is the dataset for the evaluation of the proposed method and consists of three persons' data. We set $\gamma = 0.1$ in Equation (12) and optimized the CNN model parameters by using the training dataset. The CNN is constructed by: the convolution layer on 50×1 kernel filter as the first layer, the convolution layer on 50×1 kernel filter as the second layer, the fully connected layer of five neurons as the third layer and the fully connected layer of two neurons as the fourth layer. Thus, the role of the fourth layer is the feature extraction.

First, $O_{\delta_1}(\boldsymbol{\theta})$ was used as evaluation function. The number of updating by the GD method is 500. As the result, the evaluation value $O_{\delta_1}(\boldsymbol{\theta})$ converged. We called it "optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model". Next, we switched from $O_{\delta_1}(\boldsymbol{\theta})$ to $O_{\delta_2}(\boldsymbol{\theta})$. As the result of the GD method for the optimization $O_{\delta_2}(\boldsymbol{\theta})$, the evaluation value declined many times. We stopped the updating model parameters at the 10^4 times. We called it "optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model".

The optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model and optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model are optimized by our proposal evaluation functions. To compare their models and general models of detection quality, we developed other two CNN models. One of them is non-learning model. It uses initial model parameters and we call it "initial model". The other is the CNN model optimized by cross entropy cost function. As the result of parameters updating of thirty thousand times, mini-batch error rate converged. Since it is general optimization algorithm of CNN, we called it "normal model".

After that, we developed the anomaly detection models based on the One Class-SVM (OC-SVM) in the feature spaces of the four CNN models (initial model, normal model, optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model and optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model). The kernel function of the OC-SVM we used was the Gaussian kernel. The parameter learning of the OC-SVM was performed by the normal data of the training dataset. In other words, this is the non-supervised learning.

After that, we evaluated anomaly detection quality based on F-measure. F-measure is expressed by,

$$F = \frac{2NA}{N + A}, \tag{13}$$

where N is the accuracy of the normal data detection and A is the accuracy of the anomaly data detection. The range of F-measure is from 0 to 1. If F-measure achieves high value, it means high quality detection model.

We show the results of F-measure in Table 1. Sub 1, 2 and 3 are the subjects assigned to the training dataset. Sub 4, 5 and 6 are the subjects assigned to the test dataset. In the case of initial model, F-measures are below .500 to all subjects. It means bad anomaly detection quality. In the case of normal model, anomaly detection qualities of training dataset are high. However, the values of test dataset are middle. In the case of optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model, the F-measures are similar values to all subjects and range from .612 to .782. The mean values of train and test are .718 and .699, respectively. In the case of optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model, the F-measures are similar values to all subjects and range from .685 to .837. The mean values of train and test are .776 and .761, respectively. As the results of calculating F-measure to the test dataset, it achieved a good score in order of optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model, optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model, normal model, and initial model. Therefore, the CNN optimized our proposed algorithm which can extract the features for human motions anomaly detection.

TABLE 1. F-measure of the anomaly detection models

	Training dataset			Test dataset			Mean	
Model name	Sub:1	Sub:2	Sub:3	Sub:4	Sub:5	Sub:6	Train	Test
Initial model	.142	.160	.215	.481	.390	.375	.172	.415
Normal model	.819	.880	.878	.651	.552	.642	.859	.615
Optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model	.733	.649	.772	.703	.782	.612	.718	.699
Optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model	.805	.837	.685	.767	.791	.724	.776	.761

We also survey the boundary line by the OC-SVM on the CNN-based feature space. We show the results in Figure 3 (the training data: sub 1, 2 and 3) and Figure 4 (the test data: sub 4, 5 and 6). The light gray area means the normal motions area and the dark gray area means the anomaly motions area. The white circles mean the normal motions' plot and the black circles mean the anomaly motions' plot. The left side of the figures means 0 times updating, i.e.: "initial model". The center or right sides of the figures means 500 or 10^4 times updating. These correspond to the optimization of $O_{\delta_1}(\boldsymbol{\theta})$ and optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$, respectively. In the case of not sufficient learning, the detection areas were not appropriate. In contrast, they were properly formed in the cases of the learned CNN model parameters. However, there are some anomaly plots in the normal area even in the case of learned parameters. Therefore, it is required more learning to reduce the miss-detection.

These results suggest that the CNN has acquired the robustness to human motions' individual differences by our proposal optimization algorithm because the variances of F-measure among all subjects are low in the case of optimization. However, the validation is not sufficient because the number of subjects is low. We need more verification of the reliability.

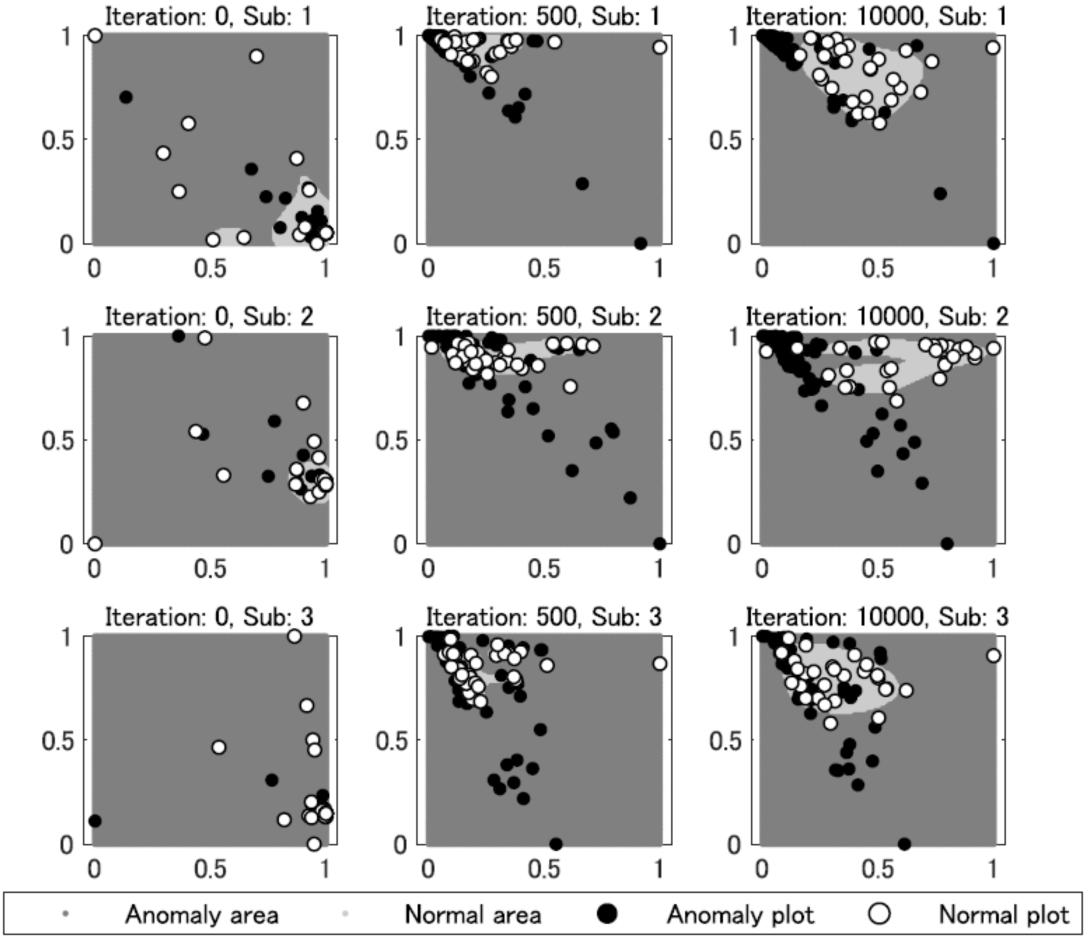

Figure 3. The boundary line of the OC-SVM (training data). The left side figures are "initial model", the center figures are "optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model" and the right side figures are "optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model".

5. **Conclusion.** In this paper, we proposed a deep learning optimization algorithm considering the individual differences of human motions from the data of inertial sensors. In our proposed algorithm, we use the Variance and Correlation (VC) evaluation function $O_{\delta_1}(\boldsymbol{\theta})$ and Consideration of Human motion's Individual differences-based Feature Space (CHI-FS) evaluation function $O_{\delta_2}(\boldsymbol{\theta})$. By using VC and CHI-FS evaluation functions, the CNN model parameters are optimized based on the Gradient Descent (GD).

We performed the experiment to confirm the reliability of our algorithm. We use the data from the swimming situation. We set the butterfly as the normal motion and the various motions (walking, jumping, leaving motion from the pool and so on) as the anomaly motions. The four CNN based-feature extraction models are (1) initial model, (2) normal model, (3) optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model and (4) optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model. As the result of comparing their F-measures, we verified that the CNN-based feature optimized $O_{\delta_{1,2}}(\boldsymbol{\theta})$ achieved most high quality as the anomaly detection models. These results suggest that by using our proposed algorithm, we can develop the human motions' anomaly detection considering the individual differences via inertial sensors based on the CNN and the OC-SVM, although the previous CNN optimization algorithms did not consider it.

As the future works, we will perform more verification of the reliability to other case studies, the reduction of the number of calculation.

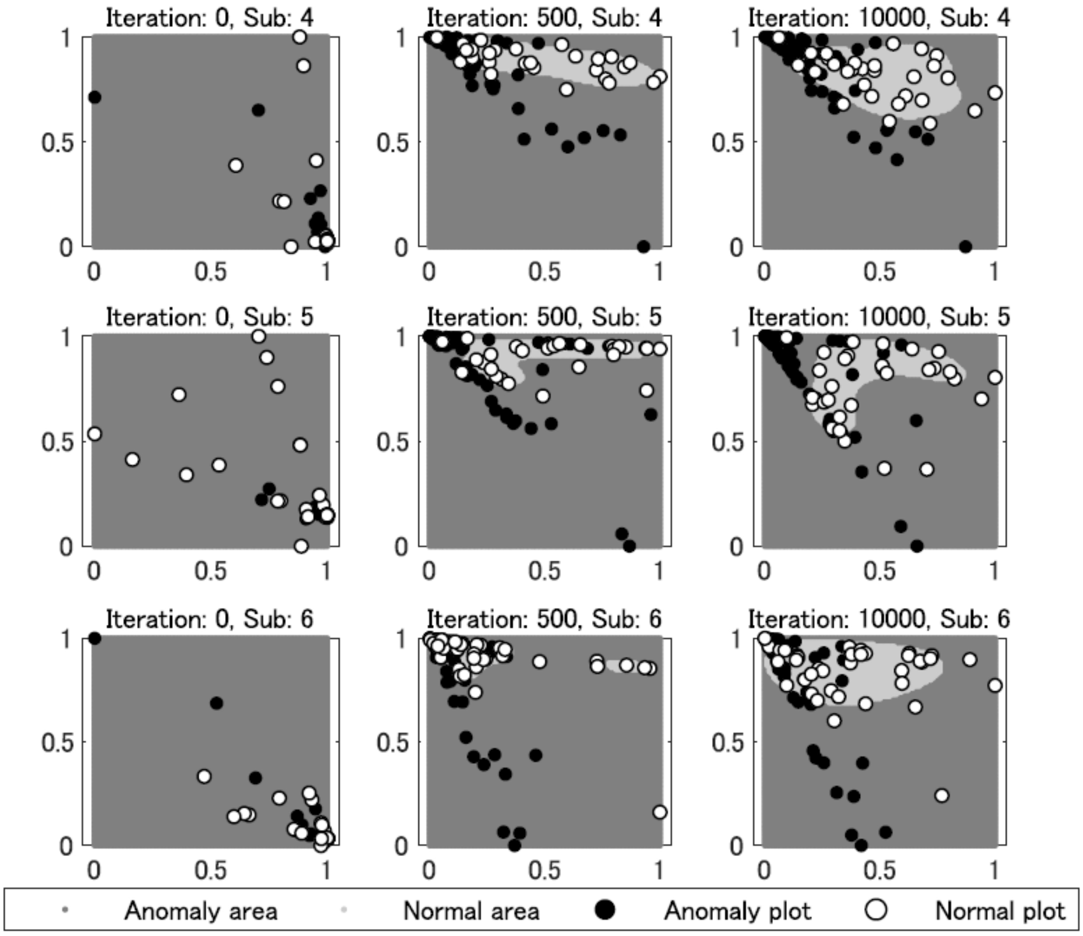

FIGURE 4. The boundary line of the OC-SVM (test data). The left side figures are "initial model", the center figures are "optimization of $O_{\delta_1}(\boldsymbol{\theta})$ model" and the right side figures are "optimization of $O_{\delta_{1,2}}(\boldsymbol{\theta})$ model".

Acknowledgment. This work was supported in part by JSPS Grant-in-Aid for Young Scientists (B) (Grant No. JP17K13179; Y. Omae) and by JSPS Grant-in-Aid for Scientific Research (C) (Grant Nos. JP16K06156 and JP17K05437; T. Akiduki and H. Takahashi).

REFERENCES

[1] D. Morris, T. S. Saponas, A. Guillory and I. Kelner, RecoFit: Using a wearable sensor to find, recognize, and count repetitive exercises, *The 32nd Annual ACM Conference on Human Factors in Computing Systems*, pp.3225-3234, 2014.

[2] L. Bao and S. S. Intille, Activity recognition from user-annotated acceleration data, *Pervasive Computing*, pp.1-17, 2004.

[3] Y. Omae, Y. Kon, M. Kobayashi, K. Sakai, A. Shionoya, H. Takahashi, T. Akiduki, K. Nakai, N. Ezaki, Y. Sakurai and C. Miyaji, Swimming style classification based on ensemble learning and adaptive feature value by using inertial measurement unit, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol.21, no.4, pp.616-631, 2017.

[4] U. Jensen, F. Prade and B. M. Eskofier, Classification of kinematic swimming data with emphasis on resource consumption, *IEEE International Conference on Body Sensor Networks*, pp.1-5, 2013.

[5] M. Kobayashi, Y. Omae, K. Sakai, A. Shionoya, H. Takahashi, T. Akiduki, K. Nakai, N. Ezaki, Y. Sakurai and C. Miyaji, Swimming motion classification for coaching system by using a sensor device, *ICIC Express Letters, Part B: Applications*, vol.9, no.3, pp.209-217, 2018.

[6] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin and H. Tan, Deep activity recognition models with triaxial accelerometers, *arXiv:1511.04664*, 2016.

[7] F. J. Ordonez and D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, *Sensors*, vol.16, no.1, p.115, 2016.

[8] D. Ravi, C. Wong, B. Lo and G. Z. Yang, A deep learning approach to on-node sensor data analytics for mobile or wearable devices, *IEEE Journal of Biomedical and Health Informatics*, vol.21, no.1, pp.56-64, 2017.

[9] Y. Omae and H. Takahashi, Feature selection algorithm considering trial and individual differences for machine learning of human activity recognition, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol.21, no.5, pp.813-824, 2017.

[10] L. Breiman, Random forests, *Machine Learning*, vol.45, no.1, pp.5-32, 2001.

[11] M. Robnik-Sikonja and I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, *Machine Learning*, vol.53, nos.1-2, pp.23-69, 2003.

[12] X. W. Chen and C. J. Jong, Minimum reference set based feature selection for small sample classifications, *The 24th International Conference on Machine Learning*, pp.153-160, 2007.

[13] M. Mori, T. Akiduki, H. Takahashi and Y. Omae, On evaluation function for feature space considered robustness to individual difference of human activities, *IEICE Technical Report*, vol.117, no.210, pp.1-6, 2017.

[14] M. Mori, T. Akiduki, H. Takahashi and Y. Omae, On reliability of feature space evaluation function based on consideration of human motions' individual differences for anomaly detection, *Proc. of the 80th National Convention of IPSJ*, 5ZA-02, 2018.

[15] M. Mori, Y. Omae, T. Akiduki and H. Takahashi, Consideration of human motion's individual differences-based feature space evaluation function for anomaly detection, *International Journal of Innovative Computing, Information and Control*, 2019 (in press).

[16] O. Banos, J. M. Galvez, M. Damas, H. Pomares and I. Rojas, Window size impact in human activity recognition, *Sensors*, vol.14, no.4, pp.6474-6499, 2014.

[17] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, CRC Press, 1986.

[18] S. J. Sheather, Density estimation, *Statistical Science*, vol.19, no.4, pp.588-597, 2004.

[19] Sports Sensing Co., LTD., *9 Axis Waterproof Type Wireless Motion Sensor*, http://www.sports-sensing.com/products/discontinued/motionwp01.html, 2018.