Augmenting Wearable Sensor Data with Physical Constraint for DNN-Based Human-Action Recognition

Hiroki OHASHI [1] **Mohammad AL-NASER** [2] **Sheraz AHMED** [2] **Takayuki AKIYAMA** [3] **Takuto SATO** [3] **Phong NGUYEN** [3] **Katsuyuki NAKAMURA** [3] **Andreas DENGEL** [2]

Abstract

A novel data augmentation method suitable for wearable sensor data is proposed. Although numerous studies have revealed the importance of the data augmentation to improve the accuracy and robustness in machine-learning tasks, the data augmentation method that is applicable to wearable sensor data have not been well studied. Unlike the conventional data augmentation methods, which are mainly developed for image and video analysis tasks, this study proposes a data augmentation method that can take an physical constraint of wearable sensors into account. The effectiveness of the proposed method was evaluated with a human-action-recognition task. The experimental results showed that the proposed method achieved better accuracy with significant difference compared to the cases where no data augmentation is applied and where a couple of simple data augmentation is applied.

1. Introduction

Human-action recognition (HAR) is important technology for many applications such as life log, healthcare, video surveillance, and worker support. A large body of works have investigated this technology in computer-vision field (Aggarwal & Cai, 1999; Turaga et al., 2008; Lavee et al., 2009; Aggarwal & Ryoo, 2011) and wearable-sensors filed (Lara & Labrador, 2013; Bulling et al., 2014; Mukhopadhyay, 2015). Recent development of deep neural-network (DNN) brought significant improvement of the performance in HAR tasks (Simonyan & Zisserman, 2014; Peng & Schmid, 2016; Ma et al., 2016). The high-learning capability of DNN gives, on one hand, much better accuracy than other methods such as support vector machine (SVM), but on the other hand it easily suffers from overfitting problem if there is not enough amount and enough variety of training data.

It is not always easy to acquire large amount and variety of training data. In order to address this issue, data augmentation methods have been employed in previous studies (Krizhevsky et al., 2012; Simonyan & Zisserman, 2015). In these examples, some of the images of an object are artificially created by changing the scale of the original image (zooming-in and zooming-out), by changing the color and brightness of the original image, and by rotating the original image. This gives the variety in the training data set, and led to the better performance.

However, the data augmentation method for wearable sensor data have not been well investigated so far. The above-mentioned augmentation methods for images are not directly applicable to the wearable sensor data simply because they have totally different characteristics including the dimension of the data.

This study proposes a novel data augmentation method that is suitable for wearable sensor data. The proposed method takes the physical constraint of the wearable sensor into account when augmenting data. As a result, it can create the data that are different from original data but still physically possible. The experimental results showed that the proposed method achieved better accuracy in a HAR task compared to the cases where no data augmentation is applied and where simple data augmentation methods such as adding Gaussian noise and transforming data without any constraint are applied.

2. Related Work and Contribution of This Study

2.1. Related Work

In the filed of HAR using wearable sensors, a lot of studies have tried various machine-learning method such as hidden Markkov model (HMM) (Guan et al., 2016), SVM (Bulling et al., 2012), conditional random field (CRF) (Adams et al.,

[1]Hitachi Europe GmbH, Germany [2]German Research center for Artificial Intelligence, Germany [3]Hitachi Ltd., Japan. Correspondence to: Hiroki OHASHI <hiroki.ohashi@hitachi-eu.com, hiroki.ohashi.uo@hitachi.com>.

ICML 2017 Time Series Workshop, Sydney, Australia, 2017.

2016), and an ensemble method (Zheng et al., 2013).

In recent years, DNN-based approaches have become increasingly popular as they showed overwhelming results in other domains such as image processing, audio processing, and natural language processing. In (Jiang & Yin, 2015; Yang et al., 2015; Ronao & Cho, 2015), researchers introduced a way to employ convolutional neural networks (CNN) to automatically extract efficient features from time-series data. Ordóñez et al. (Ordóñez & Roggen, 2016) proposed a method to more explicitly deal with the temporal dependencies of the human actions by utilizing long short-term memory (LSTM). Hammerla et al. (Hammerla et al., 2016) also introduced a LSTM-based method and gave the performance comparison among DNN, CNN, and LSTM as well as the influence of the network parameters in each method.

2.2. Contribution of This Study

One of the difficulties in HAR using wearable sensors is the displacement of the sensors. Even though a recognition model is trained using data from one subject and tested by the same subject, actions sometimes cannot be well recognized if the subject detach the sensor after collecting training data, and attach it again when testing. This happens because even if the subject tries to attach it at exactly the same position, the placement of the sensor is slightly changed and as a result the sensor readings also change accordingly.

The major way to deal with such variation of the data is to collect vast amount of training data to cover such variation. However, collecting human-action data is usually a very time consuming task. Another way to address this problem is to augment training data by applying some transformation to the original training data. Many researches in computer-vision field (Krizhevsky et al., 2012; Simonyan & Zisserman, 2015) used data augmentation and reported that it increased the accuracy and robustness.

The way of augmenting wearable sensor data, however, has not been studied well even though there have been numerous studies on HAR using wearable sensor data as reviewed in Section 2.1.

This study proposes an effective data augmentation method designed for wearable sensor data and provides a quantitative analysis on the effect of the data augmentation. This study is the first, to the best of our knowledge, to propose a data augmentation method that is suitable for wearable sensor data.

3. Proposed Method

This study uses Myo armband sensor by Thalmic Labs (Figure 1) as an example of wearable sensors. It has 8 electromyographic (EMG) sensors and inertial measurement unit (IMU) containing three-axis gyroscope, three-axis accelerometer, and three-axis magnetometer. EMG sensor measures the electric activity in muscle. This armband sensor is usually attached at a forearm.

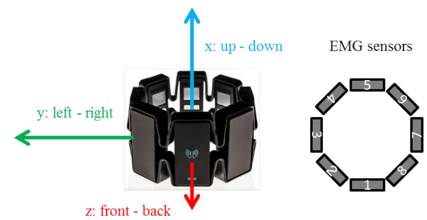

Figure 1. Myo armband sensor and its axes for IMU.

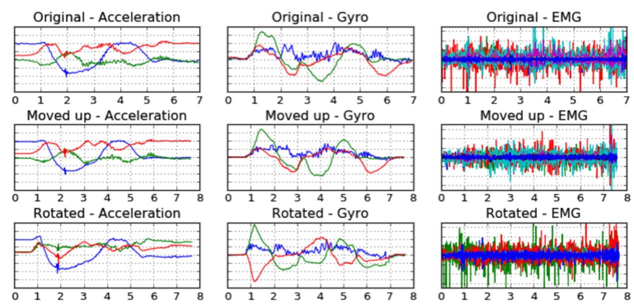

Figure 2. Impact of the sensor displacement. Moving up the sensor does not have big impact (middle), but rotating the sensor does (bottom).

When trying to recognize an action (test phase) after trained a recognition model (training phase), it is desirable to let a subject attach the armband sensor at exactly the same position on the arm as (s)he attached in the training phase. However, the slight displacement is inevitable after the subject detaches the armband sensor and attaches it again, even though (s)he tries to attach it at exactly the same position as the training phase. When the displacement happens, the displacement in up-down direction on a forearm does not have so significant impact on the sensor readings, while the rotation of the sensor does have significant impact as is shown in Figure 2. Therefore, it is desirable that the training data set includes the data collected with various rotation angles.

When the sensor is rotated, the possible rotation on an arm has some physical constraint; it can rotate only around the x-axis. Therefore, the possible rotated IMU data can be mathematically obtained by multiplying the 3-dimensional vectors of original IMU data by the following rotation matrix.

$$R_\theta = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & sin\theta \\ 0 & -sin\theta & cos\theta, \end{bmatrix} \quad (1)$$

where θ represents the rotation amount around x-axis. EMG data is measured by 8 EMG sensors shown in Figure 1. The same rotation scheme as IMU data cannot be applied to EMG data since EMG data do not have the 3-dimensional coordinate as IMU data do. Instead, the following assumption is made for virtually creating rotated EMG data. Suppose data of an arm movement are collected (original data), and data of the same movement are collected with rotated armband (rotated data), then (1) if the ith sensor (out of 8 sensors) comes to the same position on an arm as the jth sensor by the rotation, the readings of the ith sensor in the rotated data are the same as that of the jth sensor in the original data, and (2) if the ith sensor comes to somewhere between the jth sensor and $(j+1)$th sensor, the readings of the ith sensor in the rotated data can be calculated by an interpolation based on the readings of the jth and $(j+1)$th sensor in the original data and the distances from those sensors. This assumption is formulated as follows:

$$emg_i^{(\theta)} = \frac{f(d)emg_{i-\eta}^{(ori)} + f(1-d)emg_{i-\eta-1}^{(ori)}}{f(d)+f(1-d)} \quad (2)$$

$$\eta = \lfloor \theta/\phi \rfloor \quad (3)$$

$$\phi = 360/N \quad (4)$$

$$d = \theta/\phi - \eta, \quad (5)$$

where $emg_i^{(\theta)}$ denotes the EMG reading of the ith sensor in the data obtained by rotating the armband by θ degrees, $emg_i^{(ori)}$ denotes the EMG reading of the ith sensor in the original data, and N denotes the number of EMG sensors (8 in our case). Linear function ($f(d) = d$) and 2nd-order polynomial function ($f(d) = d^2$) are tested in this study.

Figure 3 shows the original data and the rotated data using the proposed rotation method (with linear function).

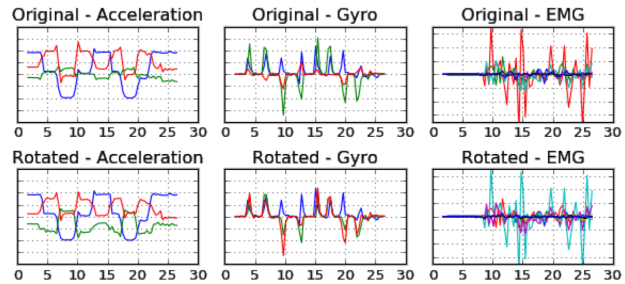

Figure 3. Original data and rotated(created) data using the proposed rotation method (with linear function). The rotation amount is 45 degrees.

4. Evaluation and Discussion

The effectiveness of the proposed method was experimentally evaluated. The target action-class in the experiment was holding, twisting, and other. For "holding", the subjects held a bag whose weight is roughly 3 kg. For "twisting", the subjects tightened a screw using a screw driver and/or opened/closed a lid of a bottle. For "other", the subjects did anything else other than "holding" and "twisting". The experimental procedure was the following. (1) a subject attaches a Myo armband sensor, (2) we start recording, (3) the subject keeps doing one action for roughly 1 minute, (4) we finish recording, (4) the subject detaches the armband (hereinafter, procedure (1) - (4) is collectively called 1 trial), (5) do (1) - (4) for the other 2 actions (hereinafter, (1) - (5), namely one trial for all the target action, is collectively called 1 series), (6) repeat (1) - (5) for several times. In each trial, subjects were asked to attach the armband at the same position as long as possible. Each subject performed at least 4 series, and in total 221 trials of data were collected (the number of trials in each action is not exactly the same since some erroneous data were removed).

In order to clarify the effect of the proposed data augmentation method, it was compared with 3 baseline methods; (1) no data augmentation (use only original data, hereinafter referred to as "original"), (2) augment data by adding Gaussian noise (hereinafter referred to as "Gaussian noise"), and (3) augment data by mathematically rotate the armband sensor without any physical constraint (not only the rotation around x-axis but also y- and z-axis are allowed, hereinafter referred to as "rotation w/o constraint"). The proposed method includes the 2 options for the augmentation of EMG data; using linear function (hereinafter referred to as "proposed (linear)") and using polynomial function (hereinafter referred to as "proposed (polynomial)"). The amount of newly created data by augmentation was same among different augmentation methods, namely 8 times more data were created by each augmentation. In "Gaussian noise" case, each trial of data were augmented by adding Gaussian noise of mean 0 and 8 different values of standard deviations ranging from 0.1ρ to 1.0ρ, where ρ denotes the standard deviation of original trial's data. In "rotation w/o constraint" case, the rotations around each 3 axis were either -15 degrees or 15 degrees and ends up with 8 (=2^3) combinations. In proposed methods, the rotation amount θ was $7.5k$ (k = -4, -3, -2, -1, 1, 2, 3, 4) degrees.

The classification model was built using DNN-based method. Statistical features and frequency-domain features were extracted by applying a sliding window and these features were fed into the network. All the parameters related to the feature extraction, network structure, and learn-

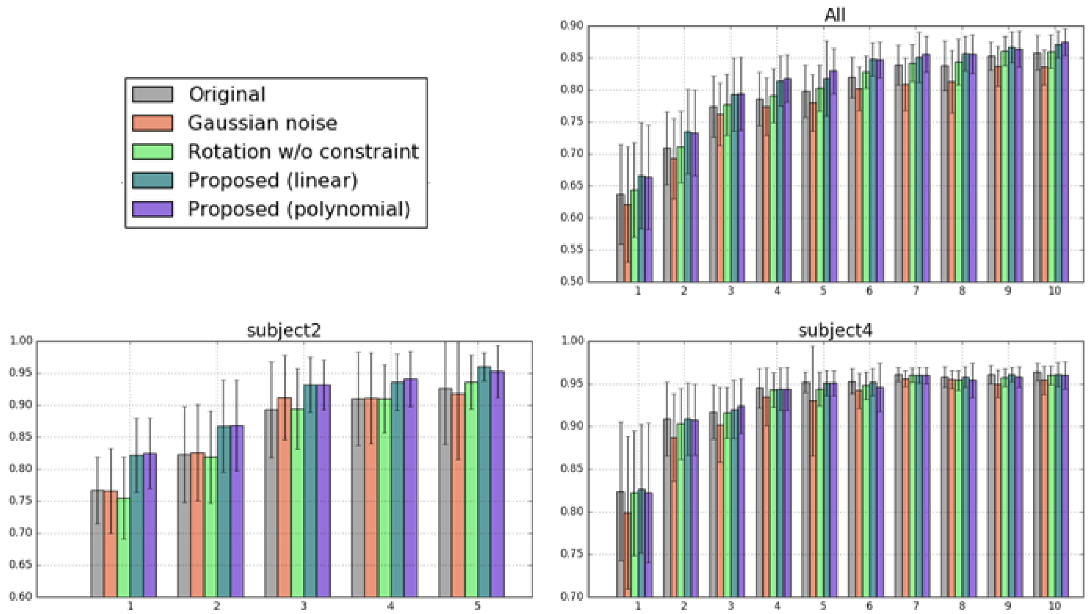

Figure 4. Test accuracies. Horizontal axis shows the number of series used as training data, while vertical axis shows the test accuracies.

ing were set to be the same among different augmentation methods (including no data augmentation case).

Both the classification accuracy for each subject (train DNN models using each subject's data and tested with the subject's data) and overall by mixing all subjects' data together were evaluated. The training data were randomly picked up from each set of data and the rest of the set were used as the test data. The number of training data picked up from each set of data were ranged from 1 series to 10 series (or (m-1) series if there are only m (<10) series in the set). For each number, the random selection of the training data and testing with the other data were performed 50 times and the average and standard deviation of them were calculated.

Figure 4 shows the test accuracies. As shown in the figure, adding Gaussian noise did not contribute to the improvement of the classification performance; actually it gave slightly worse result than "original" case. Although the sensors may suffer from small random noises, it is likely that the sensors are so stable as not to be affected by such small random noises. The augmentation by "rotation w/o constraint" gave slightly better results than "original case" in average, but the improvement does not have statistical significance. On the contrary, the proposed methods, especially "proposed (polynomial)" case did improve the accuracy compared to the original case. The t-test results on overall dataset ("All" in Figure 4) confirmed that the null hypothesis of there being no difference between the accuracy of the "original" case and that of "proposed (polynomial)" was rejected at the statistical significance level of 0.10 for the case where 1 series of data is used for training (1-series training), 0.05 for 2-series training, and 0.01 for 3- to 10- series training.

Another finding from Figure 4 is that the effect of the data augmentation is dependent on subjects. As shown in the figure, the proposed method improved the accuracy for the subject 2's data more than the other subjects. On the other hand, the impact of the data augmentation was less on the subject 4's data. The possible explanation for this difference is that while it was the first time for the subject 2 to use the armband sensor, the subject 4 had a lot of experience with the armband sensor. Therefore, the subject 4 easily found the appropriate position of the sensor on an arm and as a result the displacement of the armband sensor was relatively low.

5. Conclusion

This study has proposed a novel data augmentation method that is applicable to wearable sensor data. Unlike the conventional data augmentation methods, which are mainly developed for image and video analysis tasks, this research has proposed a method that can take an physical constraint of the wearable sensor into account. The experimental results have shown that the proposed method achieved better accuracy compared to the cases where no data augmentation is applied and where a couple of simple data augmentations are applied.

References

Adams, Roy J, Parate, Abhinav, and Marlin, Benjamin M. Hierarchical span-based conditional random fields for labeling and segmenting events in wearable sensor data streams. In *International Conference on Machine Learning (ICML)*, pp. 334–343, 2016.

Aggarwal, J and Cai, Q. Human Motion Analysis: A Review. *Computer Vision and Image Understanding*, 73 (3):428–440, 1999.

Aggarwal, J K and Ryoo, M S. Human activity analysis: A review. *ACM Computing Surveys*, 43(3):1–43, 2011.

Bulling, Andreas, Ward, Jamie A, and Gellersen, Hans. Multimodal recognition of reading activity in transit using body-worn sensors. *ACM Transactions on Applied Perception (TAP)*, 9(1):2, 2012.

Bulling, Andreas, Blanke, Ulf, and Schiele, Bernt. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR)*, 46(3): 33, 2014.

Guan, Xinze, Raich, Raviv, and Wong, Weng-Keen. Efficient multi-instance learning for activity recognition from time series data using an auto-regressive hidden markov model. In *International Conference on Machine Learning (ICML)*, pp. 2330–2339, 2016.

Hammerla, Nils Y, Halloran, Shane, and Ploetz, Thomas. Deep, convolutional, and recurrent models for human activity recognition using wearables. In *International Joint Conference on Artificial Intelligence*, pp. 1533–1540, 2016.

Jiang, Wenchao and Yin, Zhaozheng. Human activity recognition using wearable sensors by deep convolutional neural networks. In *International conference on Multimedia*, pp. 1307–1310. ACM, 2015.

Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1097–1105, 2012.

Lara, Oscar D and Labrador, Miguel A. A survey on human activity recognition using wearable sensors. *IEEE Communications Surveys and Tutorials*, 15(3):1192–1209, 2013.

Lavee, Gal, Rivlin, Ehud, and Rudzsky, Michael. Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video. *IEEE Trans. Systems, Man and Cybernetics Part C: Applications and Reviews*, 39(5):489–504, 2009.

Ma, Minghuang, Fan, Haoqi, and Kitani, Kris M. Going deeper into first-person activity recognition. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 1894–1903, 2016.

Mukhopadhyay, Subhas Chandra. Wearable sensors for human activity monitoring: A review. *IEEE sensors journal*, 15(3):1321–1330, 2015.

Ordóñez, Francisco Javier and Roggen, Daniel. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1): 115, 2016.

Peng, Xiaojiang and Schmid, Cordelia. Multi-region two-stream r-cnn for action detection. In *European Conference on Computer Vision (ECCV)*, pp. 744–759, 2016.

Ronao, Charissa Ann and Cho, Sung-Bae. Deep convolutional neural networks for human activity recognition with smartphone sensors. In *International Conference on Neural Information Processing*, pp. 46–53. Springer, 2015.

Simonyan, Karen and Zisserman, Andrew. Two-stream convolutional networks for action recognition in videos. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 568–576, 2014.

Simonyan, Karen and Zisserman, Andrew. Very deep convolutional networks for large-scale image recognition. *International Conference on Learning Representations (ICLR)*, 2015.

Turaga, Pavan, Chellappa, Rama, Subrahmanian, V. S., and Udrea, Octavian. Machine recognition of human activities: A survey. *IEEE Trans. Circuits and Systems for Video Technology*, 18(11):1473–1488, 2008.

Yang, Jianbo, Nguyen, Minh Nhut, San, Phyo Phyo, Li, Xiao Li, and Krishnaswamy, Shonali. Deep convolutional neural networks on multichannel time series for human activity recognition. In *International Joint Conference on Artificial Intelligence*, pp. 3995–4001, 2015.

Zheng, Yonglei, Wong, Weng-Keen, Guan, Xinze, and Trost, Stewart. Physical activity recognition from accelerometer data using a multi-scale ensemble method. In *Innovative Applications of Artificial Intelligence conference (IAAI)*, 2013.