Behavioral Estimation for Multiple Possession Positions Using Smartphone Accelerometers

Rui Kitahara^{1,a}, Lifeng Zhang^{1,b}

¹ Faculty of Engineering, Kyusyu Institute of Technology,
1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-8550, Japan

^a <kitahara.rui166@mail.kyutech.jp>, ^b <zhang@elcs.kyutech.ac.jp>

Abstract

With the widespread use of smartphones and wearable devices, various research has been conducted using built-in sensors. For example, height estimation and road condition estimation have been performed. In addition, behavioral estimation of the smartphone holder, possession position estimation, and person estimation has also been conducted. However, most of the measurement data is taken by fixing the possession position at a single location and not considering it in actuality when estimating behavior. In this research, we aim to estimate a person's behavior by considering multiple possession positions. It is necessary to estimate a person's behavior by considering various possession positions when using behavior estimation as a system. In addition, by treating the time series data acquired by the 3-axis acceleration sensor as a 2-dimensional image using the GAF algorithm,⁽¹⁾ class classification by machine learning is performed.

keywords: Accelerometer, Machine Learning, Gait state identification.

1. Introduction

In recent years, wearable devices such as smartphones and smartwatches have become popular. Along with this, various researches using built-in sensors have been conducted. For example, efforts are being made to estimate a person's behavior, state, and height and estimate the snow cover on the road. Another advantage of sensors is that they can protect privacy compared to data acquisition by cameras. Some systems use sensors to assist in monitoring the elderly. The following describes previous research on person behavior recognition and individual identification that is highly relevant to this research.

1.1 Related Research

1.1.1 Person identification based on acceleration sensors for multiple walking states⁽³⁾

Manabe et al. used a smartphone accelerometer to identify three types of walking conditions: level walking, stair climbing, and stair descending, and then used a discriminator defined for each walking state to identify the person. The results show that the gait state recognition rate is 95.7%. The person identification rate is 85.0% (level walking), 90.0% (stair climbing), 77.0% (stair descending) respectively. And the personal identification rate is 80.4% when performing the gait state recognition and personal identification step by step.

1.1.2 Activity Recognition using Cell Phone Accelerometers⁽⁴⁾

Jennifer R et al. acquired acceleration data using smartphones in the front pockets of 29 subjects for "walking," "jogging," "stair climbing," "sitting," and "standing" data. (WISDM dataset) The results showed that the best discrimi-

Fig. 1. Smartphone penetration rate in Japan (2021)⁽²⁾

nation accuracy was achieved using the multilayer perceptron, with 91.7%.

1.1.3 Human Activity Recognition Based on Gramian Angular Field and Deep Convolutional Neural Network⁽⁵⁾

Hongji et al. proposed a new network, Mdk-ResNet, for large-scale time-series human behavior datasets including the aforementioned WISDM dataset,⁽⁴⁾ which has higher recognition accuracy than the conventional methods. Especially for the aforementioned WISDM dataset,⁽⁴⁾ the accuracy is 9.88% higher than that of the MLP-based method by Jennifer R et al.

Although various researches have been conducted as described above, most of them are based on fixing the possession position of the smartphone in one place. However, in reality, the smartphone is not always held in one place. Therefore, the purpose of this study was to estimate the behavior of a person with a smartphone by setting multiple possession locations.

2. Principle

2.1 GAF Algorithm⁽¹⁾

The GAF algorithm⁽¹⁾ is an algorithm proposed by Zhiguang Wang for the conversion of 1D time series data into 2D image data. First, given a time series data consisting of n real-valued observation data ($X = x_1, x_2, \dots, x_n$), normalize the data so that all the values fall in the interval $[-1, 1]$ or $[0, 1]$.

$$\begin{aligned} \tilde{x}_{-1}^i &= \frac{x_i - \max(X) + (x_i - \min(X))}{\max(X) - \min(X)} \\ \text{or } \tilde{x}_0^i &= \frac{x_i - \min(X)}{\max(X) - \min(X)} \end{aligned} \quad (1)$$

Next, the normalized one-dimensional time series is transformed from a Cartesian coordinate system to a polar coordinate system.

$$\begin{cases} \phi = \arccos(\tilde{x}_i), -1 \leq \tilde{x}_i \leq 1, \tilde{x}_i \in \tilde{X} \\ r = \frac{t_i}{N}, t_i \in \mathbb{N} \end{cases} \quad (2)$$

Here, when transformed to the polar coordinate system, the value of ϕ is $[0, \pi/2]$ when normalized to $[0, 1]$, and the value of ϕ is $[0, \pi]$ when normalized to $[-1, 1]$. After transforming the normalized time series into a polar coordinate system, each temporal correlation can be identified in terms of angle by calculating the sum and difference of the trigonometric

Fig. 2. Mdk-ResNet

Fig. 3. Dilated Convolution⁽⁶⁾

functions between each sampling point. Based on the above, Gramian Summation Angular Field (GASF) and Gramian Difference Angular Field (GADF) are defined.

$$\begin{aligned} GASF &= [\cos(\phi_i + \phi_j)] \\ &= \tilde{X}' \cdot \tilde{X} - \sqrt{I - \tilde{X}^2}' \cdot \sqrt{I - \tilde{X}^2} \end{aligned} \quad (3)$$

$$\begin{aligned} GADF &= [\sin(\phi_i - \phi_j)] \\ &= \sqrt{I - \tilde{X}^2}' \cdot \tilde{X} - \tilde{X}' \cdot \sqrt{I - \tilde{X}^2} \end{aligned} \quad (4)$$

2.2 Mdk-ResNet⁽⁵⁾

Mdk-ResNet⁽⁵⁾ is a new network proposed by Hong Xii et al. for the GAF algorithm.⁽¹⁾ It boasts higher accuracy than the previous methods for WISDM,⁽⁴⁾ UCI HAR,⁽⁷⁾ and OPPOTUNITY⁽⁸⁾ time series data sets.

2.2.1 Dilated Convolution

When using large convolutional kernels, the network parameters and computational complexity increase. In addition, there is usually some information loss when using pooling layers. Therefore, Dilated Convolution solves the above problems by changing the spacing of the kernels, making it possible to increase the receptive field without information loss.

Fig. 4. Mdk-Module

2.2.2 Mdk-Res Module

The features of the previous layer are input to the Mdk-Res module, which processes them by four convolutions. The results of the four blocks and the input features are then used as the output. The hyperparameter of the convolution of the four blocks is 「padding="same",stride=1」 so that the dimensions of the output and input are the same. The green block is the previous layer of the network, the peach block has a kernel size of 1×1 , the blue block has a kernel size of 3×3 , the yellow block has a kernel size of 3×3 , dilation rate is 2, and the red block has a kernel size of 3×3 , dilation rate 4. The Mdk-Res module uses multiple normal and expanded convolution kernels simultaneously. As a result, it helps to improve the capability of the network to extract features at various scales. In addition, as with regular ResNet, residual learning is used to suppress the gradient loss problem.

3. Experiment

3.1 Data Acquisition

We set up four different positions for smartphone possession: "front pocket (outside the screen)," "back pocket (outside the screen)," "waving arms (inside the screen)," and "looking at the screen," and three different actions: "walking," "stair climbing," and "stair descending. Each data was measured for 3-axis acceleration by iPhoneSE (2nd generation). The sampling rate was set to 20Hz as in the WISDM⁽⁴⁾ data set. For five subjects, data was acquired approximately every 30 seconds per class.

3.2 Data Preprocessing

From the acquired data, the window size was set to 64 and the overlap to 50% as in the study by Hong Xii et al. The GAF algorithm was then used to convert the time series data into two-dimensional images. Since *GASF* and *GADF*

Fig. 5. Data collected by triaxial accelerometer and sliding window size

Table 1. Result

Accuracy (%)	Precision (%)	Recall (%)	F-Measure (%)
95.39	95.86	93.19	94.12

images were generated from acceleration x, acceleration y, acceleration z, six channels of images were generated per time series data.

3.3 Learning and Results

The images were converted into 2D images using the GAF algorithm,⁽¹⁾ and then trained using MDK-ResNet.⁽⁵⁾ The train:val:test ratio was set to 6 : 2 : 2. Adam was used as the optimization function, and the learning rate was set to 0.001. To prevent overtraining, Early Stopping was used. The following table shows the learning curve and the values of Accuracy, Precision, Recall, and F-Measure.

Fig. 6. Learning Curve (Accuracy)

Fig. 7. Learning Curve (Loss Function)

4. Conclusions

In this study, we evaluated the accuracy of simple action estimation and data creation considering possession position and training using Mdk-ResNet. As a result, the results were high, and we can say that it is possible to discriminate sufficiently. However, the number of data used in this study was a little small for machine learning, about 100 per class after converting to images. Therefore, in the future, we would like to increase the number of classes by increasing the number of data and adding possession positions and behaviors that were not acquired this time and verify whether it is possible to identify them.

References

- (1) Z. Wang and, T. Oates: "maging time-series to improve classification and imputation", Proc. 24th Int. Joint Conf. Artif. Intell. (IJCAI), pp. 3939-3945, Jul. 2015
- (2) Based on TCA data, 148.5%...Mobile phone penetration rate based on multiple data (latest),<http://www.garbagenews.net/archives/2059805.html>
- (3) Yusuke Manabe, and Kohji Matsuzaki : "Person identification based on acceleration sensors for multiple walking states", Intelligence and Information (Journal of the Japanese Society for Intelligent Informatics and Fuzzy Theory) Vol.27, No.5, pp.711-722,2015
- (4) Jennifer R. Kwapisz and Gary M. Weiss, "Activity Recognition using Cell Phone Accelerometers",ACM SIGKDD Explorations Newsletter,Vol12,issue2,2010
- (5) Hongji Xu, and Hui Yuan : "Human Activity Recognition Based on Gramian Angular Field and Deep Convolutional Neural Network", in IEEE Access, vol. 8, pp. 199393-199405, 2020,
- (6) F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions" in arXiv:1511.07122, 2015

- (7) D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz, "A public domain dataset for human activity recognition using smartphones", Proc. 21st Eur. Symposium Artif. Neural Netw. Comput. Intell. Mach. Learn. (ESANN), pp. 437-442, Apr. 2013
- (8) D. Roggen, A. Calatroni, M. Rossi, T. Holleczeck, K. Förster, G. Tröster, et al., "Collecting complex activity datasets in highly rich networked sensor environments", Proc. 7th Int. Conf. Netw. Sens. Syst. (INSS), pp. 233-240, Jun. 2010