Ensemble Learning for Human Activity Recognition

Ryoichi Sekiguchi
Tokyo Denki University
5 Senju-Asahi-cho, Adachi-ku
Tokyo 120-8551, Japan
20jkm17@ms.dendai.ac.jp

Kenji Abe
Tokyo Denki University
5 Senju-Asahi-cho, Adachi-ku
Tokyo 120-8551, Japan
20jkm02@ms.dendai.ac.jp

Takumi Yokoyama
Tokyo Denki University
5 Senju-Asahi-cho, Adachi-ku
Tokyo 120-8551, Japan
20jkm33@ms.dendai.ac.jp

Masayasu Kumano
Tokyo Denki University
5 Senju-Asahi-cho, Adachi-ku
Tokyo 120-8551, Japan
17aj057@ms.dendai.ac.jp

Masaki Kawakatsu
Tokyo Denki University
5 Senju-Asahi-cho, Adachi-ku
Tokyo 120-8551, Japan
kawakatu@mail.dendai.ac.jp

ABSTRACT

This paper describes a activity recognition method for Sussex-Huawei Locomotion (SHL) Challenge 2020 by team TDU_BSA. The use of ensemble learning, which combines the outputs of multiple classifiers to produce a single estimation result, improved the accuracy of activity recognition. The ensemble model consists of CNN models and a gradient-boosting model. The objective of SHL Challenge 2020 is that the users of SHL test-set are two different from SHL training-set, and the phone location of SHL test-set is not known to the SHL's participants. Therefore, estimating phone location and the user improved accuracy. SHL test-set's phone location was estimated to be Hips. The user can be estimated from SHL validation-set. The ensemble model was made with all SHL training-set (Only Hips) and 70% of SHL validation-set (Only Hips). In the submission phase, the best F-measure obtained for last 30% SHL validation-set was 84.8%.

CCS CONCEPTS

• **Computing methodologies → Activity recognition and understanding**.

KEYWORDS

Ensemble learning; CNN; Time-Frequency analysis; User estimation; Phone location estimation

ACM Reference Format:
Ryoichi Sekiguchi, Kenji Abe, Takumi Yokoyama, Masayasu Kumano, and Masaki Kawakatsu. 2020. Ensemble Learning for Human Activity Recognition. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), September 12–16, 2020, Virtual Event, Mexico.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3410530.3414346

1 INTRODUCTION

Many studies have been conducted to estimate the activity from sensors mounted on a smartphone. The authors participated in Sussex-Huawei Locomotion (SHL) Challenge 2020 as a team TDU_BSA. Our goal is to develop a activity recognition algorithm using the SHL dataset 2020 [1, 2].

In machine learning, ensemble learning that combines multiple classifiers and integrates the results to improve estimation accuracy is more effective than individual classifiers. A logistic regression model was made with the output probabilities of five classifiers as features. For train and subway, we used the results of models that were not used as input for the logistic regression model. The final estimation results were obtained.

2 SHL CHALLENGE DATA

SHL Challenge set was measured by four smartphones that were put on different positions (Bag, Hand, Hips, and Torso).

The SHL Challenge set consists of SHL training-set, SHL validation-set, and SHL test-set. SHL training-set had collected by User-1 for 59 days. Moreover, SHL validation-set had collected by User-2 and User-3 for six days. Both the datasets have four different phone positions. SHL test-set had collected by User-2 and User-3 for 40 days, but it is at one unknown location.

These datasets were sampled at 100Hz and has 9 sensors. They are acceleration sensor (Acc), gravity sensor (Gra), gyroscope sensor (Gyr), linear acceleration sensor (LAcc), geomagnetic sensor (Mag), orientation sensor (Ori) and atmospheric pressure sensor (Pre). Acc, Gra, Gyr, LAcc, and Mag has axis x, y, and z. The sensor Ori has axis x, y, z, and w.

These datasets were segmented by five seconds time-window, labeled every frame. Labels are "Still," "Walking," "Run," "Bike," "Car," "Bus," "Train," and "Subway." The objective of this challenge is to classify these data into these labels.

3 PREPROCESSING

One frame of the SHL- challenge dataset is a 5 second segment. From SHL training-set and SHL validation-set, the frames consisting of multiple activities were excluded from learning. We excluded from SHL training-set and SHL validation-set the segments that are not the same activity label from learning. The number of excluded

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8076-8/20/09…$15.00
https://doi.org/10.1145/3410530.3414346

frames was 581 in SHL training-set and 104 in SHL validation-set. We excluded these frames, but there were few change of the activity ratio. In addition, SHL training-set Hips had one frame that contained the Not a Number value in the sensor value, so it was excluded.

SHL Challenge set is sensor data in the terminal coordinate system. The values of Acc, Gyr and Mag were converted into the world coordinate system (North-East-Down Coordinate) by obtaining the rotation matrix from the orientation. Since the same axis always points up, it is possible to determine the field is coming from above or below. It is possible to reduce the influence of the difference in phone location and terminal direction. The conversion formula to the world coordinate system is shown below [3]. The rotation matrix R_{NB} was created from the quaternion $[q_w, q_x, q_y, q_z]$ and transformed.

$$R_{NB} = \begin{bmatrix} 1-2(q_y^2+q_z^2) & 2(q_xq_y-q_wq_z) & 2(q_zq_z-q_wq_y) \\ 2(q_xq_y+q_wq_z) & 1-2(q_x^2+q_z^2) & 2(q_yq_z-q_wq_x) \\ 2(q_xq_z-q_wq_y) & 2(q_yq_z+q_wq_x) & 1-2(q_x^2+q_y^2) \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_N = R_{NB} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_B$$

4 ESTIMATING USER OF SHL TEST-SET

SHL test-set is not continuous every frame. Furthermore, participants are not informed what frames are User-2 or User-3. We thought the purpose of shuffling is what participants should classify activities from a 5 second time-window. Therefore, our team did not sort SHL test-set along with time series, classified from every frame. We estimated that what frame in SHL test-set is which user, User-2 or User-3, from SHL validation-set. SHL validation-set is a part of the SHL dataset that has already distributed. So, what frame is User-2 or User-3 can be specified from sensor data in the SHL dataset.

From the SHL dataset, we specified that the first half of SHL validation-set is user-2, and the latter half is User-3. We made a CNN model for estimating which user in SHL test-set. Figure 1 is the CNN configuration. N was set to 2 because it is a user classification.

The structure of CNN is as follows. First, the combination of four sets of two convolutional layers and a max pooling layer was applied to six time-frequency spectrums. Batch-normalization and Dropout(0.2) were applied alternately with four sets of two convolutional layers and a max pooling layer. After that, they were connected by the fully connected layer and output by the soft-max function to obtain the N output probabilities. The relu function was used as the activation function on the way. We used TensorFlow's Keras API as a backend.

The used sensors were Acc, Mag, and Gyr converted to the world coordinate system. Time-frequency spectrums were created for each sensor on axis-z and the norm of a composite vector with axis-x and axis-y. The FFT time-window was 640 milliseconds (64 points), and the overlap was 80 milliseconds, and the frames were standardized along the time axis. The time-frequency spectrum was used for the features.

Considering the ratio of eight activities, we split frames of User-2, and User-3 into 7:3 along with time series. Figure 2 is a confusion matrix, that a model learned 70% of SHL validation-set, and evaluated 30% of it. Then, input SHL test-set into this CNN model, and estimated user of each frame.

5 ESTIMATING PHONE LOCATION IN SHL TEST-SET

Participants were not informed of SHL test-set phone location. Therefore, as to increase F-measure for SHL test-set, there was a need to identify SHL test-set phone location. Moreover, we attempted to identify the phone location using a time-frequency spectrum.

First, we used a gradient-boosting algorithm, the basis on a decision tree, eXtream Gradient Boosting (XGBoost), and make a model to classify eight activities. The sensors that we used are LAcc, Gyr, and Mag sensors (world coordinate system). Feature values are below. Some feature values were standardized after the calculated mean and variance of every user. The Standardization frame is noted as (Standardization).

- Mean and variance of axis-x and y norm of LAcc (Standardization)
- Mean and variance of axis-z of LAcc (Standardization)
- Skewness and kurtosis of axis-z of LAcc (Standardization)
- Sums of FFT results of LAcc axis-z splitted by every 5Hz (Standardization)
- The maximum values of FFT results of LAcc axis-z splitted by every 5Hz (Standardization)
- Frequencies when FFT results of LAcc axis-z splitted by every 5Hz get a maximum value
- Sums of FFT results of Gyr axis-z splitted by every 5Hz (Standardization)
- The maximum values of FFT results of Gyr axis-z splitted by every 5Hz (Standardization)
- Frequencies when FFT results of Gyr axis-z splitted by every 5Hz get a maximum value
- Sums of FFT results of Mag axis-z splitted by every 5Hz (Standardization)
- The maximum values of FFT results of Mag axis-z splitted by every 5Hz (Standardization)
- Frequencies when FFT results of Mag axis-z splitted by every 5Hz get a maximum value

In XGBoost, hyperparameters were adjusted, and finally the next hyperparameter was selected.

- max_depth = 16
- min_child_weight = 7
- learning_rate = 0.01
- gamma = 0.005
- sub_sample = 0.9
- colsample_bytree = 0.8
- n_estimator = 10000
- early_stopping_rounds = 30

We estimated SHL test-set's activity by the XGBoost model that learned SHL training-set and SHL validation-set. Secondly, we assumed that activities, which were "Walking," "Run," and "Car" is easy to identify the phone location. Then, the frames with an output probability of 75% or more for "Walking," "Run," and "Car" were

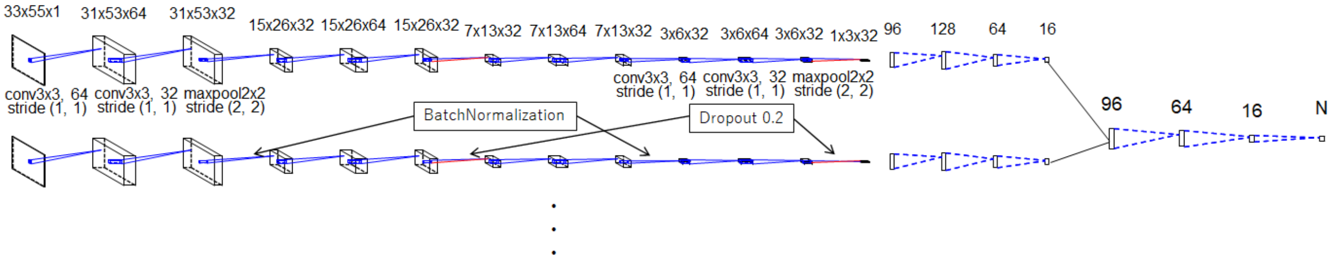


Figure 1: time-frequency CNN configuration estimating user, phone location and activity

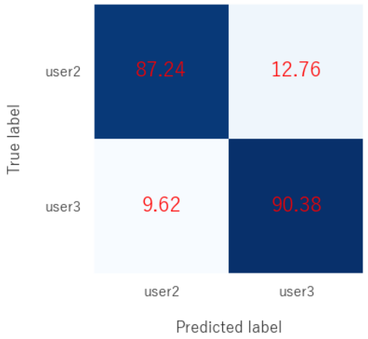


Figure 2: Confusion matrix of User estimation (SHL validation-set 30%)

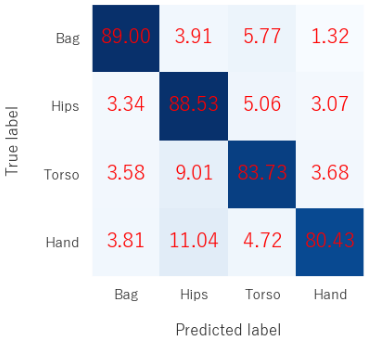


Figure 3: Confusion matrix of phone location estimation (SHL validation-set 30%)

picked up. Secondly, we built the multi-input CNN that estimates the phone location same as user estimation (Fig.1). N was set to 4 because estimated phone location (Bag, Hips, Torso, Hand).

As the feature values, the time-frequency spectrums of axis-z and the norm of a composite vector with axis-x and axis-y. The used sensors were Acc, Mag, and Gyr (terminal coordinate system). FFT time-window is 640 milliseconds, and overlap is 80 milliseconds in the time-frequency spectrum. Figure 3 is a confusion matrix for a model that learned SHL training-set and evaluated by SHL validation-set. Moreover, we inputted frames that show over 75% output probability of "Walking," "Run," or "Car" into that CNN model and estimated phone location. Then, we extracted SHL test-set frames with the maximum value of soft-max function output of 0.75 or more from the results, and checked the estimated labels. Since 79% of the extracted frames were estimated to be "Hips," our team decided to treat SHL test-set as "Hips."

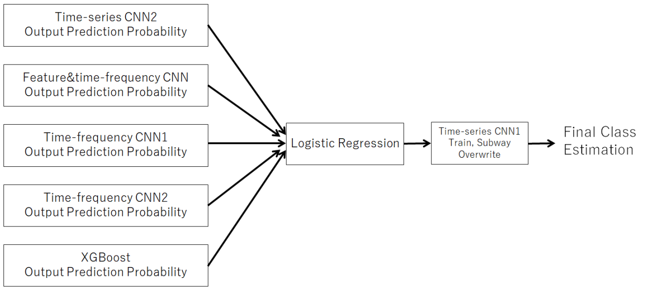

Time-series CNN2 Output Prediction Probability
Feature&time-frequency CNN Output Prediction Probability
Time-frequency CNN1 Output Prediction Probability
Time-frequency CNN2 Output Prediction Probability
XGBoost Output Prediction Probability
Logistic Regression
Time-series CNN1 Train, Subway Overwrite
Final Class Estimation

Figure 4: The flow of ensemble learning

6 METHOD

We classify activities using ensemble learning. Studies on activity recognition using ensemble learning have been conducted. Xu et al. constructed an ensemble learning model consisting of XGBoost, Random Forest, ExtraTrees, and soft-max Regression [4].

First, we created six different estimation models. The output probabilities of eight activities from six estimation models were given to logistic regression, and the final estimation result was given. SHL training-set (Only Hips) was used for training to create six different estimation models. SHL validation-set (Only Hips) was divided into 5:2:3 considering User-2 and User-3's ratio of labels in activity, respectively, and the first 50% was used for learning (our training data), and the middle 20% was used for early stopping (our validation data). The logistic regression model was made with the output probabilities of the eight activities output from the six estimation models as feature values. Figure 4 shows the flow of ensemble learning. The model used for the ensemble is as follows.

6.1 Time-Series CNN

This model was constructed using Pre and LAcc in the world coordinate system.

Figure 5 shows the flow of preprocessing. We applied with a 2 to 25 Hz band-pass filter to LAcc. As Lx and Ly, we calculated absolute values added the maximum value to the minimum value of each axis x and y. Ln was the norm of a composite vector with axis-x and axis-y. Ln2 was a feature value that we added a bigger one of Lx or Ly to Ln with each frame. About Pre, we applied with a high-pass filter through over 0.5 Hz. The features were standardized in each of SHL training-set, SHL validation-set, and SHL test-set.

The time-series CNN1 trained all of SHL training-set and our training data. For time-series CNN2, the our training data User-2 or -3 were trained from the time-series CNN1, therefore two models were created. The time-series CNN2 decided which of the

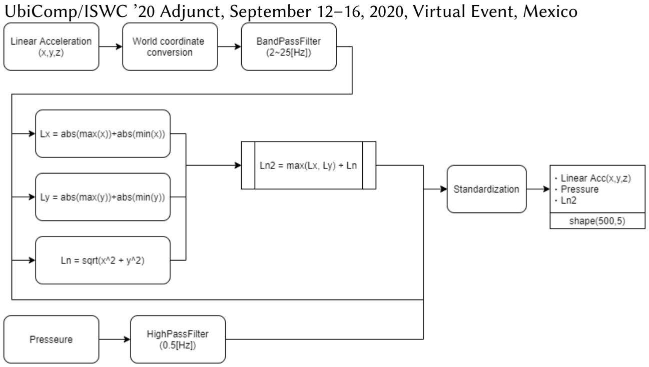


Figure 5: The flow of preprocessing

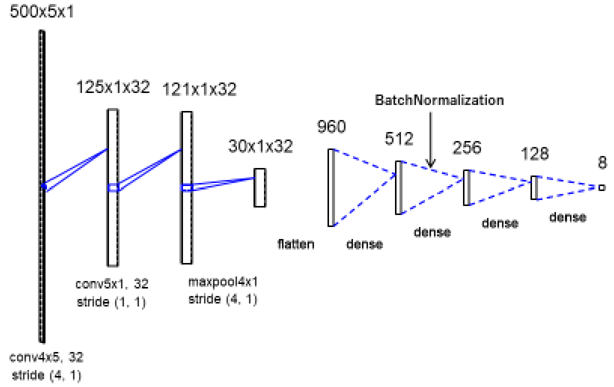


Figure 6: Time-series CNN configuration

two models to estimate the activity by using the estimation result of the user of SHL test-set.

Figure 6 is the model of a 2D convolutional neural network. The model consisted of seven layers with two convolution layers, a max pooling layer, and four fully connected layers.

6.2 Feature&time-frequency CNN

This model was constructed using LAcc and Mag in the world coordinate system. In LAcc, we made values that was sum of continuous two points of all 500 points in each frame for a axis-z and a norm of a composite vector with axis-x and axis-y. In Mag, we made time-frequency spectrums for an axis-z and a norm of a composite vector with axis-x and axis-y. The FFT time-window was two seconds, and the overlap was 100 milliseconds.

Moreover, LAcc time-series and Mag time-frequency spectrums were standardized. SHL training-set has only User-1, this was standardized in each feature value. SHL validation-set and SHL test-set has two users, User-2 and -3. In each feature value of SHL validation-set, this was standardized in each user. The SHL test-set was also standardized in data estimated as User-2 or -3.

The model had two inputs, two 1D convolution layers for the input LAcc, a 2D convolution layer for the input Mag, four fully connected layers. Figure 7 shows the CNN.

6.3 Time-Frequency CNN

This model in Figure 1 was also used for this method. Since it is a classification of activity, N was set to 8. We constructed a multi-input CNN with six time-frequency spectrums as inputs. The time-frequency spectrum used for input is the same as Chapter 4. It has been shown in past SHL Challenges that the time-frequency spectrum contributes to the activity recognition [5].

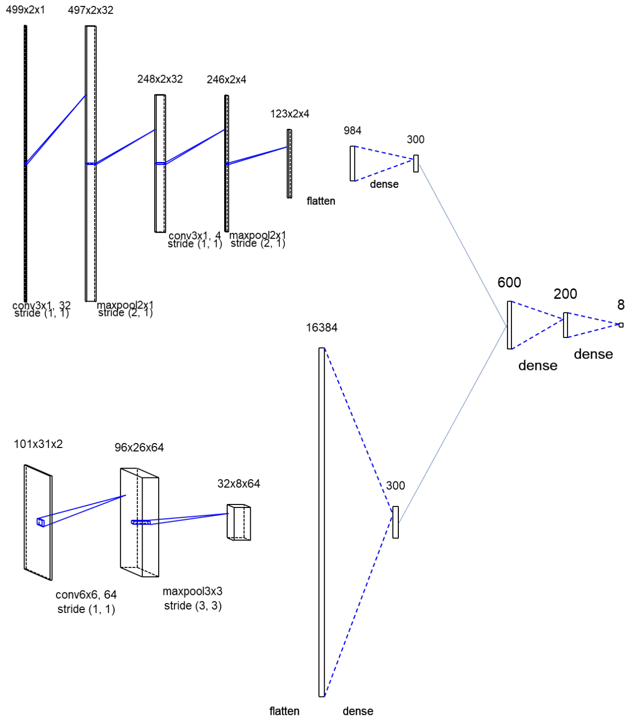


Figure 7: Feature&time-frequency CNN configuration

The time-frequency CNN1 trained all of SHL training-set and our training data. For time-frequency CNN2, our training data Users-2 or -3 were trained from time-frequency CNN1, then two models were created for user-2 and -3. Based on Chapter 4 estimation, we decided a estimation model from that two models.

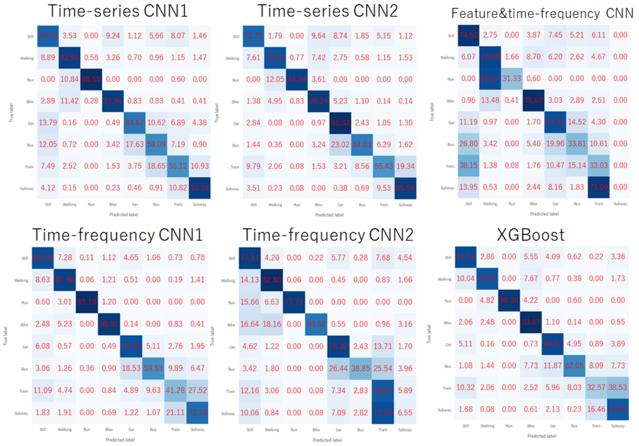
6.4 XGBoost

This model was using XGBoost for classifying eight activities. The features and hyperparameters are the same as those described in Section 5. Used sensors, as feature values are LAcc, Gyr, and Mag in the world coordinate system. Previous studies have shown that the acceleration mean, variance, skewness, and kurtosis contribute to the estimation of activity [6].

6.5 Ensemble Learning

For ensemble learning, we used time-series CNN2, Feature&time-frequency CNN, Time-frequency CNN1, 2 and XGBoost output probabilities as feature values. Our training data and our validation data were used to create the logistic regression model. SHL test-set was estimated by the logistic regression model trained by all SHL validation-set.

Next, the output labels of the logistic regression model were partially overwritten. The frame that time-series CNN1 estimated to be a train and subway was overwritten with it. Because the time-series CNN1's estimation accuracy of train and subway was higher than the ensemble models.



Time-series CNN1

	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	70.92	3.53	0.00	9.24	1.12	5.66	8.07	1.46
Walking	8.89	82.99	0.58	3.26	0.70	0.96	1.15	1.47
Run	0.00	10.84	88.55	0.00	0.00	0.00	0.60	0.00
Bike	2.89	11.42	0.28	82.94	0.83	0.83	0.41	0.41
Car	13.79	0.16	0.00	0.49	63.67	10.62	6.89	4.38
Bus	12.05	0.72	0.00	3.42	17.63	58.09	7.19	0.90
Train	7.49	2.52	0.00	1.53	3.75	18.65	55.12	10.93
Subway	4.12	0.15	0.00	0.23	0.46	0.91	10.82	83.31

Time-series CNN2

	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	71.71	1.79	0.00	9.64	8.74	1.85	5.15	1.12
Walking	7.61	78.20	0.77	7.42	2.75	0.58	1.15	1.53
Run	0.00	12.05	84.34	3.61	0.00	0.00	0.00	0.00
Bike	1.38	4.95	0.83	86.24	5.23	1.10	0.14	0.14
Car	2.84	0.08	0.00	0.97	91.32	2.43	1.05	1.30
Bus	1.44	0.36	0.00	3.24	23.02	64.03	6.29	1.62
Train	9.79	2.06	0.08	1.53	3.21	8.56	55.43	19.34
Subway	3.51	0.23	0.08	0.00	0.38	0.69	9.53	85.59

Feature&time-frequency CNN

	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	74.62	2.75	0.00	3.87	7.45	5.21	6.11	0.00
Walking	6.07	70.08	1.66	8.70	6.20	2.62	4.67	0.00
Run	0.00	68.07	31.33	0.60	0.00	0.00	0.00	0.00
Bike	0.96	13.48	0.41	76.62	3.03	2.89	2.61	0.00
Car	11.19	0.97	0.00	1.70	67.32	14.52	4.30	0.00
Bus	26.80	3.42	0.00	5.40	19.96	33.81	10.61	0.00
Train	38.15	1.38	0.08	1.76	10.47	15.14	33.03	0.00
Subway	13.95	0.53	0.00	2.44	8.16	1.83	73.09	0.00

Time-frequency CNN1

	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	84.26	7.28	0.11	1.12	4.65	1.06	0.73	0.78
Walking	8.63	87.98	0.06	1.21	0.51	0.00	0.19	1.41
Run	0.60	3.01	95.18	1.20	0.00	0.00	0.00	0.00
Bike	2.48	5.23	0.00	90.92	0.14	0.00	0.83	0.41
Car	6.08	0.57	0.00	0.49	83.05	5.11	2.76	1.95
Bus	3.06	1.26	0.36	0.90	18.53	59.53	9.89	6.47
Train	11.09	4.74	0.00	0.84	4.89	9.63	41.28	27.52
Subway	1.83	1.91	0.00	0.69	1.22	1.07	21.11	72.18

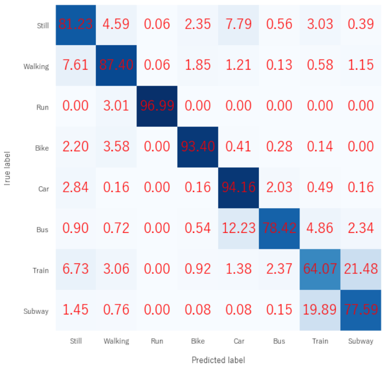
Time-frequency CNN2

	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	77.31	4.20	0.00	0.22	5.77	0.28	7.68	4.54
Walking	14.13	82.80	0.06	0.06	0.45	0.00	0.83	1.66
Run	15.66	6.63	77.71	0.00	0.00	0.00	0.00	0.00
Bike	16.64	18.16	0.00	60.52	0.55	0.00	0.96	3.16
Car	4.62	1.22	0.00	0.00	76.32	2.43	13.71	1.70
Bus	3.42	1.80	0.00	0.00	26.44	38.85	25.54	3.96
Train	12.16	3.06	0.00	0.08	7.34	2.83	68.65	5.89
Subway	10.06	0.84	0.00	0.08	7.09	2.82	72.56	6.55

XGBoost

	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	83.31	2.86	0.00	5.55	4.09	0.62	0.22	3.36
Walking	10.04	79.41	0.00	7.67	0.77	0.38	0.00	1.73
Run	0.00	4.82	90.36	4.22	0.00	0.60	0.00	0.00
Bike	2.06	2.48	0.00	93.67	1.10	0.14	0.00	0.55
Car	5.11	0.16	0.00	0.73	84.27	4.95	0.89	3.89
Bus	1.08	1.44	0.00	7.73	11.87	62.05	8.09	7.73
Train	10.32	2.06	0.00	2.52	5.96	8.03	32.57	38.53
Subway	1.68	0.08	0.00	0.61	2.13	0.23	16.46	78.81

Figure 8: Six different confusion matrices for the last 30% of SHL validation-set



	Still	Walking	Run	Bike	Car	Bus	Train	Subway
Still	81.23	4.59	0.06	2.35	7.79	0.56	3.03	0.39
Walking	7.61	87.40	0.06	1.85	1.21	0.13	0.58	1.15
Run	0.00	3.01	96.99	0.00	0.00	0.00	0.00	0.00
Bike	2.20	3.58	0.00	93.40	0.41	0.28	0.14	0.00
Car	2.84	0.16	0.00	0.16	94.16	2.03	0.49	0.16
Bus	0.90	0.72	0.00	0.54	12.23	78.42	4.86	2.34
Train	6.73	3.06	0.00	0.92	1.38	2.37	64.07	21.48
Subway	1.45	0.76	0.00	0.08	0.08	0.15	19.89	77.59

Figure 9: Confusion matrix for the last 30% of SHL validation-set

7 RESULT

Figure 8 shows the confusion matrix for the last 30% of SHL validation-set in each model. Also in this result, the label overwriting of the frame estimated by Time-series CNN1 as train and subway is applied. Figure 9 shows the confusion matrix of the ensemble learning results of the last 30% SHL validation-set. SHL test-set was estimated by the logistic regression model that also learned the last 30% SHL validation-set. Then, the frame that time-series CNN1 estimated to be a train and subway was overwritten with it. Table 1 shows the F-measure of the models.

8 CONCLUSION

In this paper, we performed ensemble learning for activity recognition. We were able to challenge the estimation of the phone location of SHL test-set and the user. We estimated that SHL test-set's phone location was Hips, then we made six different models. The features for logistic regression model were output probabilities of Time-series CNN2, Feature&time-frequency CNN, Time-frequency CNN 1 and 2, and XGBoost. But in "Train" and "Subway", we used the estimation results of Time-series CNN1. In the final verification, a F-measure of 84.8% was obtained for last 30% SHL validation-set. From this result, it is expected that high accuracy will be obtained for the estimation of the activities of SHL test-set. The recognition result for the testing dataset will be presented in the summary paper of the challenge [7].

Table 1: The F-measure of the models (SHL validation-set 30%)

Model name	F-measure
Time-series CNN1	72.5%
Time-series CNN2	76.0%
Feature&time-frequency CNN	46.1%
Time-frequency CNN1	76.5%
Time-frequency CNN2	62.0%
XGBoost	74.5%
Logistic regression	83.9%
Logistic regression (Overwrite)	84.8%

Table 2: The detail of the models created in this challenge

Model name	Model size	Learning time
Time-series CNN	5.4 MB	2 hours
Feature&time-frequency CNN	21.1 MB	2 hours
Time-frequency CNN	6.9 MB	8 hours
XGBoost	80.5 MB	20 minutes

COMPUTER RESOURCES

CPU core i7, 3.19GHz, 6 core, 12 threads, RAM 32GB, Geforce RTX 2060 super. Table 2 shows the learning time of the models created in this challenge.

REFERENCES

[1] H. Gjoreski, M. Ciliberto, L. Wang, F. J. Ordonez Morales, S. Mekki, S. Valentin, and D. Roggen. The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access*, 6:42592–42604, 2018.

[2] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, and D. Roggen. Enabling reproducible research in sensor-based transportation mode recognition with the sussex-huawei dataset. *IEEE Access*, 7:10870–10891, 2019.

[3] Vito Janko, Mitja Lustrek, Nina Reščič, Miha Mlakar, Vid Drobnič, Matjaz Gams, Gašper Slapničar, Martin Gjoreski, Jani Bizjak, and Matej Marinko. A new frontier for activity recognition: The sussex-huawei locomotion challenge. pages 1511–1520, 10 2018.

[4] Shoujiang Xu, Qingfeng Tang, Linpeng Jin, and Zhigeng Pan. A cascade ensemble learning model for human activity recognition with smartphones. *Sensors*, 19(10):2307, May 2019.

[5] Chihiro Ito, Masaki Shuzo, and Eisaku Maeda. Cnn for human activity recognition on small datasets of acceleration and gyro sensors using transfer learning. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, UbiComp/ISWC '19 Adjunct, page 724–729, New York, NY, USA, 2019. Association for Computing Machinery.

[6] Keisuke Yoshida, Shogo Matsuno, and Minoru Ohyama. Highly accurate method for estimating movement state of smartphone users. In *Communications and Signal Processing 2016*. RISP International Workshop on Nonlinear Circuits, 03 2016.

[7] L. Wang, H. Gjoreski, M. Ciliberto, P. Lago, K. Murao, T. Okita, and D. Roggen. Summary of the sussex-huawei locomotion-transportation recognition challenge 2020. Proceedings of the 2020 ACM International Joint Conference and 2020 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 2020.