Let's not make it complicated - Using only LightGBM and Naive Bayes for macro and micro activity recognition from a small dataset

Ryoichi Kojima, Roberto Legaspi, Kiyohito Yoshihara, Shinya Wada

Abstract We propose a model that combines only simple techniques to meet the challenge of cooking activity recognition. The challenge dataset is basically small, consisting only of four subjects where three are used for training and one for validation. In order not to overfit the small training data, we employed two simple classifiers, LightGBM and Naive Bayes, which suited the task. To prevent leakage from other subject data during training, we used Leave One Subject Out cross validation. Further, we incorporated a postprocessing step wherein the Naive Bayes corrects the macro activity classification outcomes that have been derived by LightGBM, based on the combinations of macro and micro activities that are likely to occur. We hypothesized that this added postprocessing will improve the macro actvitiy recognition, and with it our model may be able to adapt well and generalize to other small datasets. As a result, our proposed model achieved an average accuracy of 0.557 when classifying macro and micro activities from a small dataset.

1 Introduction

The spread of accessible and easy to use sensors has increased significantly the possibility of obtaining massive amount of data. Together with this is the development of machine learning algorithms that have made remarkable progress in human activity

Ryoichi Kojima
KDDI Research, Tokyo Japan, e-mail: ry-kojima@kddi-research.jp

Roberto Legaspi
KDDI Research, Tokyo Japan, e-mail: ro-legaspi@kddi-research.jp

Kiyohito Yoshihara
KDDI Research, Tokyo Japan, e-mail: yosshy@kddi-research.jp

Shinya Wada
KDDI Research, Tokyo Japan, e-mail: sh-wada@kddi-research.jp

recognition research [1][2][3][4][5]. The kinds of data that are handled in this research area vary from common activities, such as walking and sitting, to specialized ones like nursing care [8] and dangerous activities like falling down [10].

It is often the case, however, that open datasets for specialized human activity recognition research (such as the ones cited above) are small, which is problematic for complex machine learning tasks. For instance, recent works on computer vision-based human activity recognition place premium on formulating complex deep learning techniques. Comes along with it, however, is the requirement (or should we say curse) to obtain and process large training datasets, which is usually very costly. In other words, although recent breakthroughs from deep learning techniques have demonstrated that advanced algorithms and complex architectures can endow human-like intelligence to machines, the caveat is that they are data-hungry - large amount of training data is critical for them to be successful.

A machine learning model that does not overfit a given small dataset and performs optimally the classification task is therefore required. The dataset [6][7] from the pertinent Cooking Activity Recognition Challenge [9] is basically small, recorded from only four subject participants, thereby posing the same compelling problem mentioned above. Our approach so as not to overfit during training is to employ simple classifiers, specifically, LightGBM for macro and micro activity recognition, and Naive Bayes to correct at a postprocessing stage the macro activity classification outcomes derived by LightGBM. Indeed, we are cognizant that we are not introducing a novel classifier, but we argue for the value of this two-level processing model.

We organized the paper as follows. We describe in Section 2 the Cooking Activity Recognition Challenge and its dataset. We then elaborate in Section 3 the manner by which we conducted the preprocessing and feature engineering. Section 4 details our classification model. We then elucidate in Section 5 the outcomes of our validation experiments and the viability of our model given the results. We explain in Section 6 how our work is situated with other research works in the domain of human activity recognition. Finally, we conclude in Section 7. In addition, we describe in the Appendix the computer resources and processing time consequent of our work.

2 Challenge and Dataset Details

The purpose of this challenge is to formulate a highly accurate model that recognizes macro and micro cooking activities from the dataset [6][7]. A macro activity is a cooking recipe, while a micro activity is a low-level activity that is a component of a macro activity. Table 1 shows the count of macro activity files for each subject used for training the model, and Table 2 shows the number of micro activities included in each macro activity. We note here that certain micro activities do not exist for some macro activities.

The dataset used for training consists of three subjects, while the test dataset came from another subject. Each subject was instructed to prepare three recipes, namely, sandwich, fruit salad and cereal, for five times each. Although the subjects

Table 1 Count of macro activity files for each subject used during training

	Subject 1	Subject 2	Subject 3
sandwich	39	38	36
fruit salad	16	44	42
cereal	25	23	25

Table 2 Count of micro activities per macro activity

	#Add	#Cut	#Mix	#Open	#Peel	#Pour	#Put	#Take	#Wash	#Other
sandwich	0	34	0	0	0	0	47	43	30	49
fruit salad	18	41	19	0	69	0	41	50	0	14
cereal	0	24	0	23	27	23	26	41	0	22

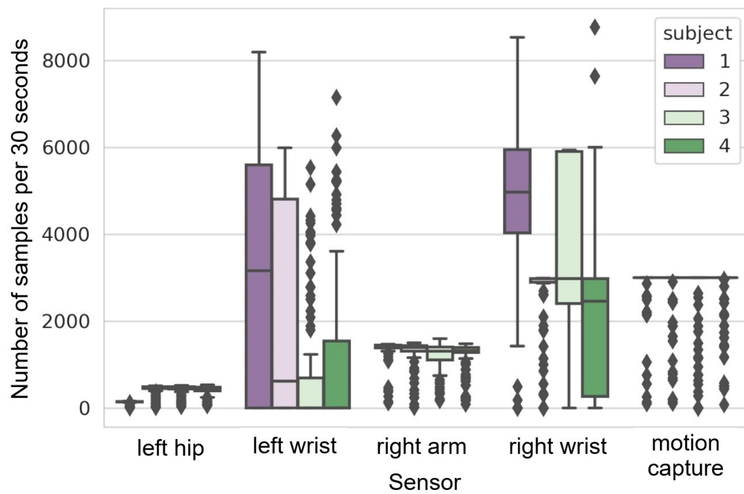

Fig. 1 Varying sampling rates. It can be observed that the number of samples collected from each sensor differ. Furthermore, the motion capture system shows greater stability in recording data. In contrast, the accelerometer data have missing samples, especially those recorded by the left and right wrist accelerometer sensors.

followed a script for every recipe, they acted as naturally as possible. To record the subjects' gestures, which were then used to label the activities, two smartphones (placed on the right arm and left hip), two smartwatches (on both wrists) and one motion capture system with 29 markers were used. When we looked at the sampling rates used to collect the data (shown by the box-plot in Fig.1), we observed the sampling rates varied between sensors. Further, we observed the motion capture system demonstrated more stability in recording data in contrast to the accelerometer records that have missing samples. Hence, we needed to perform missing data interpolation to obtain uniform sampling and approximate the missing data.

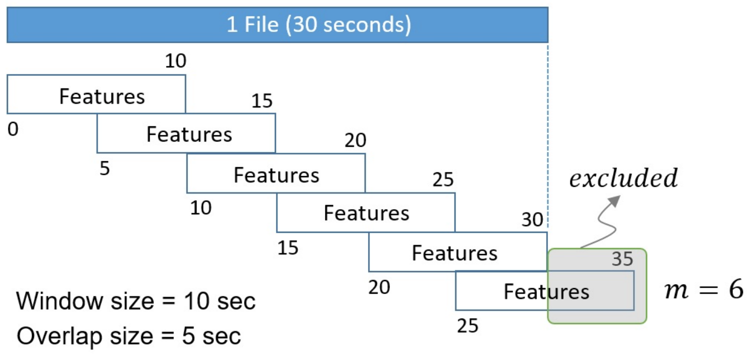

Fig. 2 Feature engineering. To generate the features, we segmented each 30-second file into m blocks (see Eq. (1) for the computation) using an overlapping sliding window. We experimented on different combinations of (window, overlap) sizes. Specifically, we tried the following combinations in unit seconds: (10,1), (10,5), (5,3), (5,1), (15,10), (15,5), (15,3) and (15,1). As shown in the image, we excluded the portion of the feature block that exceeded 30 seconds. The overlapping sliding window of size 10 seconds with overlap of 5 seconds proved optimal.

The data collected from each subject is divided into 30-second files and were randomly shuffled. For each file, macro activity recognition becomes a multi-class classification task, and micro activity recognition becomes a multi-label classification task. The estimated result is evaluated by the average accuracy of the macro and micro activity recognition tasks.

3 Preprocessing and Feature Engineering

During preprocessing, we simulated a uniform sampling of the accelerometer and motion capture sensor data for a sampling rate of 10Hz. Naturally, this generated additional missing values. While we fill all missing values in the accelerometer data with 0s, we performed linear interpolation to estimate the missing motion capture data. We assumed the missing values in the accelerometer data means that the subjects did not move, hence, we augmented with 0s. On the other hand, since the motion capture system sent out absolute coordinates, linear interpolation is appropriate.

We engineered the sensor data features using an overlapping sliding window over the 30-second file, shown in Fig.2. The window and overlap (in unit seconds) are hyperparameters and therefore need to be tuned. Depending on the window and overlap sizes, and we tried various combinations, the file is divided into m blocks:

$$m = \frac{30\,sec}{(\text{window size} - \text{overlap size})}. \tag{1}$$

If a portion of a block exceeds the duration of the file, shown highlighted in Fig.2, the excess part is excluded from the computation of features.

The subsequent feature engineering methods differed for the accelerometer and motion capture data. We extracted statistical features from the accelerometer data, specifically, the mean, standard deviation, maximum and minimum values, from each feature block. As for the motion capture data, since their coordinates are different, it became necessary for us to extract features that are different from the accelerometer's. We achieved this by employing the features used by Eusha Kadir and colleagues [11] to recognize nursing activities from motion capture markers placed on the upper body. What is common with cooking and nursing activities is that the participants stay at a specific place and mainly use their upper body to perform their tasks.

Specifically, for the 11 marker points, which are the Top.Head, Rear.Head, R.Shoulder, R.Offset, R.Elbow, R.Wrist, L.Shoulder, L.Elbow, L.Wrist, R.ASIS and L.ASIS, we generated the following features from each feature block (window):

- distance between barycentric coordinates of the markers; and
- direction vector from one barycentric coordinate of a marker to another.

Consequently, all in all, we generated a total of 110 features.

4 Classification Model

We classified both macro and micro activities using LightGBM [12]. LightGBM, albeit a simple classifier, has been successfully applied to human activity recognition (e.g., refer to [13] on how the use of LightGBM outperforms other methods). Although we tried simpler classification methods, e.g., k-nearest neighbor, their performance were unstable with the classification accuracy varying greatly between subjects. We tuned the parameters using Optuna [14], which is an open source optimization software that includes a define-by-run API to dynamically construct the parameter search space and automates the hyperparameter search. We trained the LightGBM using Leave One Subject Out cross validation (LOSOCV), i.e., two out of the three subjects are used to build the training data, and the remaining one subject is used for the validation data. The objective in using LOSOCV is to prevent leakage from other subjects during training.

We recognized the macro activities into three classes using one LightGBM. Specifically, we defined the macro activities making sandwich, fruit salad and cereal to be M_1, M_2 and M_3, respectively, and we denote the corresponding prediction probabilities learned by LightGBM to be

$$p_l(M_1), p_l(M_2), p_l(M_3). \tag{2}$$

On the other hand, we used binary classification for the micro activities using 10 LightGBMs. To achieve this, first, we constructed a matrix L that contains the prediction probabilities learned by every LightGBM for each feature block of every micro activity. We constructed L as

$$L = \begin{pmatrix} l_{1,1} & l_{1,2} & \dots & l_{1,m} \\ l_{2,1} & l_{2,2} & \dots & l_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ l_{10,1} & l_{10,2} & \dots & l_{10,m} \end{pmatrix}, \tag{3}$$

where $l_{i,k}$ is a prediction probability for micro activity i and feature block k: with $i = 1..10$ corresponding to the 10 micro activities, and $k = 1..m$ with m, see Eq. (1), denoting the number of feature blocks in the 30-second file (in Fig.2). We then generated a matrix L' to contain the mean of each of the rows in L, which is

$$L' = \frac{1}{m} \begin{pmatrix} \sum_{k=1}^{m} l_{1,k} \\ \sum_{k=1}^{m} l_{2,k} \\ \vdots \\ \sum_{k=1}^{m} l_{10,k} \end{pmatrix} \coloneqq \begin{pmatrix} l'_1 \\ l'_2 \\ \vdots \\ l'_{10} \end{pmatrix}. \tag{4}$$

Further, we compared each l' in L' to a threshold t to generate the binary classification, that is,

$$\sigma(l') = \begin{cases} 1 & (l' \geq t) \\ 0 & (l' < t), \end{cases} \tag{5}$$

where the threshold t is determined by maximizing the prediction accuracy of our model for each micro activity in the validation data. Hence, the value of t varied depending on the micro activity. Using now σ, the micro activity binary classification results are given by

$$\begin{pmatrix} \sigma(l'_1) \\ \sigma(l'_2) \\ \vdots \\ \sigma(l'_{10}) \end{pmatrix}. \tag{6}$$

Finally, from Table 2, we can observe the co-occurrences of macro and micro activities. This motivated us to add a postprocessing step, i.e., to re-estimate with the Naive Bayes classifier the macro activities, this time utilizing the micro activities predicted by the LightGBMs. Hence, the postprocess probabilities of each macro activity is given by

$$\begin{aligned} p_n(M_i) &\coloneqq p(M_i | \sigma(l'_1), \sigma(l'_2), \dots, \sigma(l'_{10})) \\ &= \frac{p(M_i) * \prod_{j=1}^{10} p(\sigma(l'_k) | M_i)}{p(\sigma(l'_1), \sigma(l'_2), \dots, \sigma(l'_{10}))}, \quad (i = 1, 2, 3). \end{aligned} \tag{7}$$

From Eqs. (2) and (7), the macro activity can now be re-estimated as

$$\arg\max_{M_i} \frac{p_l(M_i) + p_n(M_i)}{2}, \quad (i = 1, 2, 3). \tag{8}$$

5 Experimental Results and Discussion

Table 3 shows the performance of our proposed model when classifying the cooking activities in terms of the mean classification accuracy for macro activities (ma), mean accuracy of micro activity classification (mi), and their average ($(ma+mi)/2$). Table 3 differentiates the activity recognition accuracies of our proposed model when the LightGBMs alone are used and when their micro activity predictions are used further to inform the Naive Bayes classifier for postprocessing estimations.

Again, our hypothesis is that the postprocessing stage with the Naive Bayes that takes into account the co-occurrences of macro and micro activities will improve the recognition of macro activities. Table 3 shows that our hypothesis is indeed correct, with the mean recognition accuracy for macro activities improving from 0.454 to notably 0.569, and the mean recognition accuracy considering both macro and micro activities improving from 0.500 to about 0.557.

We now discuss how our proposed method can address three difficult issues. The first is how to achieve high generalization accuracy from a small data set. This also means not to overfit the training data. We solved this rather simply by using LOSOCV that does not put the same subject in the training and validation dataset at the same time. The more accurate the classifier becomes, the more we need to be careful in ensuring that leakage does not happen between datasets. Training without including the same subject in the train and validation datasets enables more robust classification from unseen subjects.

The second is that it is unclear in the datasets how many micro activities are actually included within the classification range of 30 seconds. To go around this problem, we created a model that uses 10 binary classifiers, each one corresponding to each of the 10 micro activities. Here, the 10 binary classifiers have different thresholds to determine each activity label. Had we used a single multi-label classifier with only one threshold, it would be difficult to classify activities whose numbers are unknown.

The third one is that the distribution of macro and micro activities is not independent of each other. Specifically, there is a large bias in the distribution of micro activities contained in the macro activities (refer to Table 2). Therefore, the ability of our model to recognize correctly the activities can be improved by correcting the macro activity classification, which we achieved by using Naive Bayes (in Table 3).

Table 3 Accuracy of our model in recognizing cooking activities

Method	ma	mi	(ma+mi)/2
LightGBM	0.454	0.546	0.500
Postprocessed LightGBM using Naive Bayes	**0.569**	0.546	**0.557**

6 Background and Related Works

One of the challenging goals of machine learning research is to make accurate classification of human behaviours. Human activity recognition has been extensively studied, demonstrated for example by well-cited literature surveys that have been written on this subject in various domains and applications, and employing different methods that have proved to achieve promising results (e.g., surveys from [1], [15], [2] and [16], with newer ones still flowing in like [17] and [5], among others). Human activity recognition has been used for a number of applications, including health-care services and smart homes. Many sensors have been utilized for this purpose, which include wearables, smartphones, RF-based and motion capture systems.

The current state of the art methods in human activity recognition use advanced algorithms based on deep learning, which proved to achieve great success in other challenging research areas as well, e.g., image recognition and natural language processing to say the least. Until a few years ago, classical discriminant models such as k-nearest neighbor and logistic regression were mainly used. Recently, however, deep learning models have been employed since they can automate feature engineering to a good extent, hence, it is not required to handcraft the features. Since human activity recognition uses data with different characteristics, there seems to be an unending stream of newly proposed methods to this end (e.g., [18], [19], [20], and others). But automated feature engineering is not the focus of our contribution.

The other key merit of deep learning is learning automatically the representative features from massive data. We reckon, however, that this is also the curse of deep learning methods - being data-hungry, the use of small data becomes detrimental to deep learning methods that leads to their poor recognition accuracy. We hypothesized that when confronted with a small amount of data, we do not need to resort to advanced yet complex methods and employ instead simple classifiers.

7 Conclusion

We propound simple classifiers are sufficient for recognizing macro and micro activities from a cooking challenge dataset with high accuracy. Recent deep learning models have high accuracy, but contingent with a large amount of data to work with. In other words, training is suboptimal at best and the accuracy is poor if the amount of data is small. On the other hand, since our proposed method is simple, sufficient training is possible even with small data. Further, postprocessing based on the co-occurrence behavior of macro and micro cooking activities improved the classification accuracy of our model. We hypothesized that with this added postprocessing, our method may be able to adapt well and generalize to other small datasets. Hence, our future work is to estimate human activity from other datasets using the method we elucidated here.

Appendix

We summarize in Table 4 the proposed method we elucidated in this paper, as well as the experiment environment in which our method was carried out.

Table 4 Summary of our proposed model and experiment environment

	Details
Used sensor modalities	Accelerometer and motion capture sensors
Accelerometer features	Mean, standard deviation, max and min values
Motion capture features	Distance between barycentric coordinates of the markers, and direction vector from one barycentric coordinate of a marker to other one
Programming language	Python 3.6
Libraries	sklearn 0.22.2, pandas 1.0.3, lightgbm 2.3.2, optuna 1.0
Classifiers	LightGBM, Naive Bayes (postprocessing)
Parameters	Automatically tuned by Optuna [14]
Validation scheme	LOSOCV
Tested (window, overlap) sizes	(10,1), (10,5), (5,3), (5,1), (15,10), (15,5), (15,3) and (15,1), all in unit seconds
Optimal window size	10 seconds
Optimal overlap size	5 seconds
Training time	530 seconds
Testing time	20 seconds
Machine specification	CPU : 4.00GHz, RAM : 32GB, GPU : no

References

1. Vrigkas, M., Nikou, C., Kakadiaris, I.A.: A review of human activity recognition methods. Front. Robot. AI **2**, 28 (2015). DOI 10.3389/frobt.2015.00028. https://www.frontiersin.org/article/10.3389/frobt.2015.00028
2. Zhang, S., Wei, Z., Nie, J., Huang, L., Wang, S., Li, Z.: A review on human activity recognition using vision-based method. Journal of Healthcare Engineering **2017**, 31 (2017). DOI 10.1155/2017/3090343. https://doi.org/10.1155/2017/3090343
3. Twomey, N., Diethe, T., Fafoutis, X., Elsts, A., McConville, R., Flach, P., Craddock, I.: A comprehensive study of activity recognition using accelerometers. Informatics **5**(2), 1–37 (2018). DOI 10.3390/informatics5020027. https://www.mdpi.com/2227-9709/5/2/27/htm
4. Wang, J., Chen, Y., Hao, S., Peng, X., Hu, L.: Deep learning for sensor-based activity recognition: A Survey. Pattern Recognition Letters **119**(1), 3–11 (2019). DOI 10.1016/j.patrec.2018.02.010. https://www.sciencedirect.com/science/article/abs/pii/S016786551830045X
5. Jobanputra, C., Bavishi, J., Doshi, N.: Human activity recognition: A survey. Procedia Computer Science **155**, 698–703 (2019). DOI 10.1016/j.procs.2019.08.100. http://www.sciencedirect.com/science/article/pii/S1877050919310166
6. Lago, P., Takeda, S., Adachi, K., Alia, S.S., Matsuki, M., Benaissa, B., Inoue, S., Charpillet, C.: Cooking activity dataset with macro and micro activities. IEEE DataPort (2020). DOI 10.21227/hyzg-9m49

7. Lago, P., Takeda, S., Alia, S.S., Adachi, K., Benaissa, B., Charpillet, F., Inoue, S.: A dataset for complex activity recognition with micro and macro activities in a cooking scenario. Preprint (2020)
8. Lago, P., Alia, S.S., Takeda, S., Mairittha, T., Mairittha, N., Faiz, F., Nishimura, Y., Adachi, K., Okita, T., Charpillet, F., Inoue, S.: Nurse care activity recognition challenge: Summary and results. UbiComp/ISWC 2019 - 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, pp. 746–751 (2019). DOI 10.1145/3341162.3345577. https://dl.acm.org/doi/10.1145/3341162.3345577
9. Alia, S.S., Lago, P., Takeda, S., Adachi, K., Benaissa, B., Ahad, M.A.R., Inoue, S.: Summary of the Cooking Activity Recognition Challenge. Human Activity Recognition Challenge, Smart Innovation, Systems and Technologies (2020), Springer Nature.
10. Martínez-Villaseñor, L., Ponce, H., Brieva, J., Moya-Albor, E., Núñez-Martínez, J., Peñafort-Asturiano, C.: Up-fall detection dataset: A multimodal approach. Sensors (Basel, Switzerland) **19**(9) (2019). DOI 10.3390/s19091988. https://www.mdpi.com/1424-8220/19/9/1988
11. Eusha Kadir, M., Akash, P.S., Sharmin, S., Ali, A.A., Shoyaib, M.: Can a simple approach identify complex nurse care activity? UbiComp/ISWC 2019 - 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, pp. 736–740 (2019). DOI 10.1145/3341162.3344859. https://dl.acm.org/doi/10.1145/3341162.3344859
12. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems **30**, 3146–3154 (2017). https://github.com/Microsoft/LightGBM.
13. Gao, X., Luo, H.,Wang, Q., Zhao, F.,Ye, L., Zhang,Y.: A human activity recognition algorithm based on stacking denoising Autoencoder and LightGBM. Sensors (Basel, Switzerland) **19** (2019). DOI 10.3390/s19040947
14. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2623–2631 (2019). DOI 10.1145/3292500.3330701. https://dl.acm.org/doi/10.1145/3292500.3330701
15. Ziaeefard, M., Bergevin, R.: Semantic human activity recognition: A literature review. Pattern Recognition **48**(8), 2329-2345 (2015). DOI 10.1016/j.patcog.2015.03.006. https://www.sciencedirect.com/science/article/abs/pii/S0031320315000953
16. Ramasamy Ramamurthy, S., Roy, N.: Recent trends in machine learning for human activity recognition - A survey. WIREs Data Mining and Knowledge Discovery **8**:e1254 (2018). DOI 10.1002/widm.1254. https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1254
17. Ye, J., Dobson, S., Zambonelli, F.: Lifelong learning in sensor-based human activity recognition. IEEE Pervasive Computing **18**(3), 49-58 (2019). DOI 10.1109/MPRV.2019.2913933. https://ieeexplore.ieee.org/document/8903481
18. Ordóñez, F.J., Roggen, D.: Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors **16**(1), 115 (2016). DOI 10.3390/s16010115. https://www.mdpi.com/1424-8220/16/1/115
19. Yao, S., Hu, S., Zhao, Y., Zhang, A., Abdelzaher, T.: DeepSense: A unified deep learning framework for time-series mobile sensing data processing. Proceedings of the 26th International World Wide Web Conference, pp. 351–360 (2017). DOI 10.1145/3038912.3052577. https://dl.acm.org/doi/10.1145/3038912.3052577
20. Ma, H., Li, W., Zhang, X., Gao, S., Lu, S.: AttnSense: Multi-level attention mechanism for multimodal human activity recognition. Proceedings of the 28th International Joint Conference on Artificial Intelligence, pp. 3109–3115 (2019). DOI 10.24963/ijcai.2019/431. https://www.ijcai.org/Proceedings/2019/431