An Energy-Efficient Human Activity Recognition System Based on Smartphones

Junhao Shi

Harbin Institute of Technology

Decheng Zuo (✉ zuode@ftcl.hit.edu.cn)

Harbin Institute of Technology <https://orcid.org/0000-0002-8386-2858>

Zhan Zhang

Harbin Institute of Technology

Daohua Pan

Harbin Institute of Technology

Research

Keywords: Human activity recognition, Energy-efficient, Deep learning, Pervasive computing, Smartphone application

Posted Date: October 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-89249/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

An energy-efficient human activity recognition system based on smartphones

Shi Junhao¹, Zuo Decheng^{1,*}, Zhang Zhan¹, Pan Daohua¹

Abstract:

Smartphone-based human activity recognition has become a considerable research field as a subdomains of pattern recognition and pervasive computing. With the increasing popularity of smartphones, HAR has prominent applications in number of fields such as health care, education, entertainment and etc. Smart devices have a huge advantage in convenience as the main acquisition and processing equipment, but the battery life of smartphone and other resources are limited for long-duration tasks. In this paper, we propose a lightweight HAR system. The system realizes HAR algorithm with deep learning algorithm. Beyond that, we introduce a clustering-center based pre-classification strategy to reduce the call frequency of the DL model. Meanwhile, we add a sampling frequency control mechanism to the inertial sensor. The goal of the whole system is to achieve low power consumption and time delay. According to the final experiment results, the energy consumption reduces about 49% and time delay reduces about 55% while the overall recognition accuracy only suffers about 10% reduction.

Key word: Human activity recognition; Energy-efficient; Deep learning; Pervasive computing; Smartphone application

1 Introduction

Human activity recognition (HAR) is a vital research filed in pervasive computing and also provides the background of various applications. By using internal sensors of smart phones, researchers can collect daily information in areas such as healthcare, entertainment, education, etc. [1-3]. For example, an application is specifically developed for fall-detection to help the elderly in [4]. A medical monitoring system integrated multiple sensors is designed to detect physiological data in [5]. Most of these HAR researches focus on improving the recognition accuracy under the specific application background.

In recent years, inertial sensors in phones have become a primary HAR data source with the increasing popularity of smart devices. Wang et al. [6] build a cloud platform using smart phone as the core device, and recognize activities (including running, riding bike, driving, etc.) of students

¹ Department of Computer Science and Technology, Harbin Institute of Technology, Heilongjiang, China

* Corresponding author

This work described in this paper was supported by the National High-tech R&D Program of China (863 program) under Grant 2013AA01A215

More author information available on the last page of the article.

to evaluate psychological state and academic performance of the users by gathering sorts of sensing data. Generally, smartphone-based HAR is a long-duration sensing task while the power of smart phones is limited, which makes energy-efficient necessary for activity recognition. In a system involving multiple types of sensors, it is usually necessary to turn off some sensors (such as GPS) with high energy consumption to save power [7], while in some single sensor activity recognition systems, the energy saving is realized by reducing sampling frequency of the sensor. For example, Morillo et al. [8] use a constantly changing data sampling frequency from 32Hz to 50Hz to collect data and identify activities (including walking, jumping, cycling, etc.). As mentioned above, energy-saving methods inevitably reduce the accuracy of HAR. Therefore, the core of the problem lies in the balance between low power consumption and recognition accuracy (or other performance indexes).

In our work, we proposed a lightweight HAR system for daily activities. The mainframe is based on deep learning (DL) algorithm for activity recognition, and pre-classification strategies are introduced to save energy. The overall work is detailed as follows:

First, we utilize the accelerometer in the mobile phone to form an activity dataset of multiple subjects. Meanwhile, a public dataset is also utilized in our work. Based on fully convolutional network (FCN), which is a classic DL network, we conduct cyclic training and testing with the previous dataset to gain the best network structure and parameter settings through fine-tuning. In order to show the advantages of DL, we introduce 4 machine learning classifiers which are commonly used in similar pattern recognition systems for comparison, including Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN) [9-12]. Moreover, we also introduce another 2 works using DL models. In [13], Chi Yoon Jeong et al. designed a deep learning model based on fully convolutional network structure, which is proved to be energy-efficient. In [14], Pienaar et al. proposed a Long Short-Term Memory (LSTM) based classifier.

Second, we define the concept of cluster-center is also proposed for pre-classification. For all activities in the dataset, the cluster-center is calculated respectively, which is regarded as the core feature of an activity. The detailed information is depicted in the following parts. Meanwhile, we conduct experiments to explore the influence of sensor sampling frequency on recognition accuracy. Finally, we transplant the whole system on Android platform and deploy it on smartphone. By testing on new subjects, we record the classification accuracy, power consumption and time delay of the proposed system to certificate the energy saving effect.

We mainly make three contributions in our work: 1) An activity recognition model based on DL, which is a simplified fully convolutional network (SFCN), is trained by fine-tuning parameters, and the model is transplanted to the Android platform for real scenario test. 2) We introduce the pre-classification strategy to reduce the energy consumption and time delay of our HAR system, we call the whole model as pre-classification-based deep learning model (PC-DL). 3) We explore the relationship between sampling frequency and recognition accuracy through experiments, and determine the most appropriate sampling frequency for the purpose of power control.

The rest of this paper is organized as follows: In part 2, we introduce related work about our research. Part 3 depicts details about each module of our HAR system. Part 4 shows the experiment results about model training and testing and parameter configuration of our pre-classification strategies. Finally, conclusion and future work is given in Part 5.

2 Related Work

2.1 Sensor-based activity recognition

Sensor-based human activity recognition has been studied for years. Some early researches make particular sensing device for specific recognition targets. Pansiot et al. [15] develop the e-ar sensor, which can be worn on the ear to detect human body signs data for health care. Minnen et al. [16] put multiple sensors on military suit to recognize tactical actions and provide battlefield information. Angelini et al. [17] design a smart bracelet for the elderly. With the popularity of smart phones, many studies on activity recognition take advantage of smart phones as the core device to collect, process and identify data. Akhavian et al. [18] attach mobile phone to the upper arm of construction workers in their study to collect work data for identifying the activities of workers. Tran et al. [19] deploy the SVM model on a smartphone to recognize the daily activities of the human body. Bisio et al. [20] utilize sensors integrated in smart phones for medical monitoring on patients. Smart phone is a potential pervasive computing platform due to its convenience and popularity. However, as a device for long-duration tasks like HAR, its computing and energy resources are limited.

2.2 Deep learning algorithm

DL algorithms have attracted much attention these years and make excellent performance in image processing, NLP (natural language processing), data mining and other fields [21-23]. Some classical network structures are designed such as AlexNet [24], Resnet [25], etc. The features extracted by the DL model are believed to be able to reveal deep character of the data. DL model is also applied to the research of sensor-based activity recognition recently. Yang et al. [26] utilized CNN network structure to the recognition of multi-channel human activity data sequence, and the results were superior to the traditional classifier. Hammerla et al. [27] compared performance of various DL network on HAR. In general, DL method has better recognition performance and robust characteristics, while it consumes more computing and memory resource than traditional machine learning algorithm, which limits its availability on smart phones [28]. Due to the advantage of LSTM in dealing with time series issues, several efforts focus on using LSTM for HAR. T. Zebin et al. [29] present a Long-Short Term Memory (LSTM) deep recurrent neural network for the classification of six daily life activities from accelerometer and gyroscope data. Wang et al. [30] propose a Hierarchical LSTM architecture for activity recognition.

2.3 Energy-efficient strategy

The existing energy-efficient strategy for based HAR realizes low power consumption by adjusting sensor sampling frequency. For example, Reddy et al. [31] change sampling frequency via detecting the user location information using GPS, they achieve an overall recognition accuracy of 93.6% with data from 16 subjects, but GPS consumes huge power and is not suitable for indoor tasks. Luis et al. [32] discretize origin data to reduce the computation load, and experiments on smart phones showed that the battery life is extended to 27 hours, but the minimum sampling frequency still reached to 32 Hz. Yan et al. [33] propose an adaptive method, which established relationship between sampling frequency and feature set, and achieved 20%-25% energy consumption reduction. Qi et al. [34] design the AdaSense model, and prove that multiple activity recognition consumed more resource than judging a single activity is in progress or not, hence the algorithm increases the time proportion of single activity detection, and reduces the energy consumption by 39.4%.

The above energy-efficient strategies only reduce power consumption through adjusting sampling frequency and feature subset selection, without consideration on the classifier optimization.

Meanwhile, there is no discussion on time delay of the proposed systems, which is crucial to wearable applications especially like epilepsy detection, fall detection, etc. [35]. So our work attempt to propose a pre-classification method based on the clustering-center, which can be seen as a simple core feature for HAR

Also, some energy efficient works presuppose more complex scenarios. In [36], David Chu et al. propose a tool Kobe aids mobile classifier development, and kinds of scenarios and sensors are involved. In [37], the author weights sensors to activities and synthesizes the context information for energy saving. Comparing with these works, our target activity set is designed simpler. This is based on the consideration that the systems may be utilized in recognizing military tactical operations, which are relatively simple and standard. The background of our work is quite different from the above research, still we benefit from these works.

3 System architecture

3.1 System main frame

We develop an Android app for HAR, the architecture is shown in Fig. 1.

Figure 1 System architecture

Our system architecture mainly consists of two parts. One is offline training module, which is in charge of training DL model and pre-classification. The other is online recognition module. In this module, the system utilizes pre-classification model to obtain primary recognition results. These results can be considered as final results if they are acceptable through judgment, otherwise the DL model will be activated for recognition and final results are fed back to cluster-center calculating as correction reference. Meanwhile, the activity fluctuation is also calculated, and sampling frequency is real-time controlled according to the computing result.

3.2 Data acquisition and preprocessing

In our work, we utilize two dataset to conduct the experiments. First, we introduce a public dataset—WISDM [38] as a benchmark. This dataset (v 1.1) contains accelerometer data of 6 activities (including Walking, Jogging, Upstairs, Downstairs, Sitting, Standing) collected from 36 subjects. The sampling frequency is 20 Hz (1 sample every 50ms). This dataset has been adopted in amount of studies such as Reference [39], therefore it is representative in the HAR filed.

Moreover, we build a HAR data set using the accelerometer in a MI phone on the right thigh of subject to train the model and here we simply name it as self-collected dataset (SCD). The

experiment invites 8 subjects (6 male and 2 female, age 22-27) to perform 8 activities, including sitting, standing, lying on the left side, upstairs, downstairs, walking, running, quick walk. For each activity of a subject, we gather data at a frequency of 50Hz for 3 minutes. A 3-order low-pass Butterworth filter is applied to eliminate the noise, then the data is segmented with sliding window algorithm. According to previous experimental experience, we define the window size as 2s, which is close to the activity duration [40].

3.3 DL-HAR model

AlexNet is a DL model designed by Hinton and his team in 2012, which is one of classical DL models. It includes 5 convolutional layers and 3 full-connected layers, and each layer contains thousands or even more neurons. Given the size of our HAR tasks, we simplify AlexNet by cycling fine-tuning to form our DL-HAR model. The final net structure is given in Table 1

Table 1

Structure of FCN and SFCN

Layer	FCN	SFCN
1	Input(50*3)	Input(50*3)
2	Conv2D 16 filter *(3*3)	Conv2D 16 filter *(3*3)
3	Conv2D 16 filter *(3*3)	Conv2D 16 filter *(3*3)
4	Max Pooling (2*2)	Max Pooling (2*2)
5	Conv2D 32 filter *(3*3)	Conv2D 16 filter *(3*3)
6	Max Pooling (2*2)	GAP
7	Conv2D 6 filter *(3*3)	Softmax
8	GAP	
9	Softmax	

Refer to [41], we use a full-connected convolutional net (FCN) structure, which replace the full-connected layers with global average pooling (GAP) layers. The GAP contains less parameters that can lighten the net structure. Furthermore, we design another simplified-FCN (SFCN) in our work to further reduce the energy consumption. The final structure preserves 3 convolutional layer and 1 GAP layer. The segmented data blocks are sent to the DL model for recognition. We use $\mathbf{X}$ to represent the input data. Fig 2 gives the process of convolution. After convolution, feature maps are produced. The element $E_{p,q}^k$ of the kth feature map can be represented as:

$$E_{p,q}^k = F(b^k + \sum_{(\alpha,\beta)=(1,1)}^{(i,j)} \omega_{\alpha,\beta}^k X_{\alpha,\beta}^{p,q}) \quad (1)$$

Where p and q represent the location of $E_{p,q}^k$ in the feature map. b^k is the bias of the feature map. Convolution kernel size is expressed as $i * j$. $\omega_{\alpha,\beta}^k$ represents the weight of the convolution kernel and (α, β) is filter index. $\mathbf{X}^{p,q}$ represents the sensing data matrix with its first element locates in the position (p, q) . F is the activation function, which can be denoted as follow formula:

$$F(x) = \max(0, x) \quad (2)$$

After the convolution, all the feature maps are changed to 1D vector after the GAP layers as follow formula:

$$GAP(M) = \text{average}(m_1, m_2, \dots, m_n) \quad (3)$$

Where M depict a feature map and $m_i \in M$ ($1 \leq i \leq n$).

Figure 2 Process of convolution

The recognition results are obtained in form of probability after the softmax layer.

We then realize the whole net and train it. The forward propagation is performed and error values are calculated after every epoch. Meanwhile, weight update and error cost minimization is performed using the Root Mean Square Prop (RMSProp) algorithm on minibatches of the data examples. More information about the training stage is given in Section 4.1.

3.4 Pre-classification strategy

We propose two pre-classification strategies to improve the overall performance of the model in this section. We define the concept of clustering-center for pre-classification. DL model performs excellent on pattern recognition issues with high energy and other resources consumption. The proposed pre-classification strategy in our work is to save energy by reducing the call frequency of DL model.

In the HAR field, a feature vector consists of time-domain and frequency-domain features extracted from activity data for recognition. Frequency-domain features perform well on periodic activities while the computational complexity and delay are relatively large than time-domain features which are more commonly used. Therefore, our system only takes advantage of time-domain features and the recursive feature elimination (RFE) algorithm is applied to reduce the dimension of feature vector to 36. (Feature 1-11 are extracted from the accelerometer data of 3-axis and Feature 12 is the coefficient of 3-axis data as shown in Table 2).

Table 2

Features of Activity data

No.	Feature	Formula
1	Mean	$\bar{a}$
2	Variance	$\sum_{i=1}^n (a_i - \mu)^2$
3	Maximum	$\max(a_i)$
4	Minimum	$\min(a_i)$
5	Range	$\max(a_i) - \min(a_i)$
6	ZCR	$\sum_{i=1}^n \text{sig}(a_i > 0)$
7	Median	$\text{median}(a_i)$
8	MAD	$\text{median}(a_i - \text{median}(a_i))$
9	Information Entropy	$-\sum_{i=1}^m (p_i * \log(p_i))$
10	Kurtosis	$E[(a_i - \mu)^4 / \sigma^4]$
11	Skewness	$E[(a_i - \mu)^3 / \sigma^3]$
12	Coefficient	$\text{cov}(X, Y)$

*ZCR: Zero crossing rate MAD: Absolute median difference

In the offline training stage, feature vectors of all data blocks from same activity are calculated, and then we define the clustering-center as the clustering value of all feature vectors of an activity. We denote the clustering-center set of all activities as $T = \{T_i | i \in \{1, 2, \dots, c\}\}$, where c represents the number of activities to be classified. The feature center can be calculated as:

$$T_i = \text{argmin}_{V_{ij}} \forall_{j \in N_i} \sum_{k \in N_i} \text{dist}(V_{ij}, V_{ik}) \quad (4)$$

Where V_{ij} is the feature vector extracted from the j th training data of i th activity (denote as A_i), $N_i = \{1, 2, \dots, n_i\}$, where n_i is the training size of A_i , the dist is the distance between two vectors and depicted as Equation (5):

$$\text{dist}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

In the online recognition stage, for current time window t , X_t is the time domain feature of t , and calculate the distance between the feature center T_i and X_t , denoted as $\text{dist}(X_t, T_i)$, and the $E = (e_1, e_2, \dots, e_c)$ represents the Euclidean distance between elements of T and the to-be-classified feature vector. e_i reflects the possibility whether the current activity in the current time window is A_i : the greater e_i is, the more X_t deviates the feature center of A_i , and so the lower the possibility that the user is carrying out A_i and vice versa.

For the vector E above, there may be some cases that little difference between e_i and e_j , which means pre-classification strategy is unable to recognize t . Consider of this, we define a constraint to determine whether accepting the pre-classification results. The formula is as follow:

$$\min(\{e_1, e_2, \dots, e_c\}) \leq \text{conf}_i \quad (6)$$

Where conf_i is the pre-classification confidence of A_i and calculated as follow:

$$\text{conf}_i = \text{Quantile}_\beta(\{\text{dist}(V_{ij}, T_i) | j \in N_i\}) \quad (7)$$

In (7), the function Quantile_β gets the β -quantile of a sequence. The experiment of determining β is depicted in Part 4. If (7) is considered true, the pre-classification result will be taken as the final recognition result. For each A_i , conf_i is a constant after being calculated in offline training stage.

3.5 Adaptive frequency control strategy

The complexity of activities differs from each other. Obviously, static activities like sitting and lying are significantly less complex than dynamic activities include running and walking. In the

study, we believe that activity recognition with low complexity also requires less data or features compare with high complex activity, which is the main theoretical basis to reduce the sampling frequency to save energy. Therefore, we propose an adaptive sampling frequency control strategy, which adjust sampling frequency referring to the history record, to realize energy-efficient recognition.

We introduce the concept Activity Changing Rate (ACR) to measure the probability of the user staying in the current activity. The formula is shown as follow:

$$\sigma_t = \sqrt{\frac{1}{\lambda} \sum_{i=1}^{\lambda} (stren_i - \overline{stren})^2} \quad (8)$$

Where t represents the current time window and $stren_i$ is the activity intensity of A_i . The concept of activity intensity is a constant and based on the value of energy consumed by the activity [34] as shown in Table 3. Since the sampling frequency of WISDM is fixed, here we use the SCD data, which is in multi sampling frequency, as a benchmark.

Table 3

Activity intensity of 8 activities

Activity	Sitting	Standing	Lying	Upstairs
Intensity	1.0	2.0	1.0	8.0
Activity	Downstairs	Walking	Running	Quickwalk
Intensity	3.0	2.5	10.0	4.0

σ_t is the standard deviation of the activity intensity calculated from the past λ time windows, which is just the ACR we defined. When σ_t is smaller than threshold represented as δ , the activity state is considered stable which means the sampling frequency can be reduced. Obviously, the threshold δ has a decisive influence on the system sampling frequency. On the other hand, the parameter λ represents the quantity of referred historical time windows when calculating σ_t . It is objective using long enough historical data when recognizing activity, while large data lead to more storage and computation resource consumption. We determine the value of δ and λ through experiments in Section 4.

4 Experiment and result

4.1 DL model training and classifier comparison

The DL model in our system stems from the structure of AlexNet through fine-tuning. Fig. 3 shows the optimization process:

(a)

(b)

Figure 3 Relation between DL model recognition accuracy and training epochs (a) and kernel size (b)

In the net training, we find multiple convolutional layers have no positive effect on the overall accuracy. Too many layers may lead to the degradation of the effect, which was consisted with many related studies [35]. Since the data blocks were 100*3, the 3*3 kernel size was suitable for the system. 5 training epochs were appropriate and more epochs had little effect on accuracy.

To see the recognition ability of DL model, we take several other traditional machine learning model and DL model for comparison. First, we compare the recognition accuracy of various classifiers through experiments. We divide the WISDM dataset into 10 sections and utilize 10-fold cross validation on these 7 models (SVM, RF, DT, KNN, FCN, LSTM, and SFCN). The results are shown in Fig 4, where the A01-A06 represent “Walking”, “Jogging”, “Sitting”, “Standing”, “Upstairs”, “Downstairs”.

Figure 4 Confusion Matrix of 7 classifiers (WISDM)

According to the results in Fig 4, all the classifiers performs well on A01-A04, with an accuracy rate close to 100%. However, the FCN, LSTM has obvious advantages over traditional machine learning classifiers on recognizing A05 and A06, and the recognition accuracy are all above 80%. The overall results of SFCN is almost the same as that of FCN, which indicates that deep structure may not always improve the performance of the model. In addition, we also take time delay as another important parameter into consideration when choosing the model. We believe that time delay and energy consumption are positively correlated, so its value can also partly reflect the energy consumption degree. We deploy the experiment on the computer with a configuration as Table 4 shows.

Table 4
Configuration of the PC platform

Attribute	Value
CPU	Inter i5-7500 CPU 3.40GHz
RAM	16GB
OS	Windows 10

We run the experiment for 10 times and take average time delay per sample of all the 5 classifiers as the final results, which are given in Fig 5.

Figure 5 Time delay of the classifiers

It is obvious that DL algorithms such as FCN, LSTM and SFCN have high latency due to their complex structures. The time consumption is between 1 ms and 1.2 ms per sample. In contrast, the time delay of traditional machine learning algorithms is below 0.7 ms per sample.

In order to test the generalization ability of the models, we conduct a leave-one-out validation. In the WISDM dataset, we find 36 subjects totally. Due to the lack of a small amount of data, we finally utilize 32 of them for the experiment. The results are shown in Table 5.

Table 5
Leave-one-out Validation Result

Subject No.	SVM	RF	DT	KNN	FCN	LSTM	SFCN
33	0.9158	0.9443	0.7501	0.9049	0.9389	0.9171	0.9145
17	0.5731	0.4541	0.3896	0.4323	0.6412	0.3143	0.6921
29	0.7909	0.8254	0.7997	0.7763	0.8991	0.8071	0.9148
13	0.8879	0.9186	0.8124	0.8927	0.9763	0.9704	0.9768
20	0.8616	0.8457	0.8125	0.8397	0.7475	0.6871	0.8824
27	0.7872	0.6808	0.5755	0.6922	0.6693	0.7065	0.6997
6	0.8755	0.9125	0.7860	0.8466	0.8957	0.8849	0.8917
15	0.9482	0.9283	0.8629	0.9176	0.8814	0.8352	0.8587
32	0.8699	0.8958	0.8592	0.8778	0.8925	0.9065	0.8739
36	0.8327	0.8730	0.5985	0.7720	0.9232	0.9102	0.8717
18	0.8903	0.8927	0.8221	0.8934	0.7672	0.5911	0.8123
11	0.6153	0.5578	0.4201	0.5559	0.6519	0.6519	0.7284

16	0.7041	0.7051	0.6768	0.7480	0.9277	0.8672	0.9473
5	0.7791	0.8401	0.8275	0.7433	0.9060	0.9502	0.9180
10	0.7754	0.8046	0.8115	0.7966	0.9049	0.9095	0.8865
28	0.9054	0.8689	0.7912	0.8399	0.9288	0.9579	0.9560
26	0.8011	0.8398	0.7238	0.8189	0.9227	0.8336	0.8882
14	0.7817	0.8272	0.6874	0.7857	0.8330	0.8634	0.9010
24	0.6674	0.7071	0.6154	0.7529	0.7451	0.7312	0.6193
12	0.6871	0.6240	0.6277	0.6803	0.6785	0.7103	0.7373
23	0.8474	0.8326	0.7196	0.7469	0.8053	0.8769	0.8450
4	0.7399	0.8295	0.6292	0.6766	0.7188	0.7522	0.6942
30	0.1744	0.1800	0.2052	0.1626	0.2289	0.1752	0.1807
34	0.8565	0.8487	0.7625	0.7570	0.8954	0.7419	0.8671
8	0.8864	0.8689	0.7311	0.8345	0.9612	0.9058	0.9228
31	0.8772	0.8700	0.7666	0.7856	0.8677	0.8957	0.9174
21	0.8753	0.8572	0.7602	0.7630	0.8933	0.9137	0.9001
3	0.8442	0.8107	0.6157	0.7789	0.8908	0.8954	0.8874
22	0.7258	0.7330	0.4758	0.5815	0.8781	0.7867	0.8593
1	0.7045	0.7225	0.7285	0.7312	0.8245	0.7618	0.8252
7	0.8709	0.8754	0.7624	0.8541	0.9496	0.8005	0.8677
19	0.8493	0.8291	0.7700	0.8352	0.9317	0.8604	0.9278
Average	0.7875	0.7876	0.6930	0.7523	0.8305	0.7929	0.8332

As the results show, the average accuracy of FCN, LSTM and SFCN is significantly higher than that of other models. It suggests that the DL methods has better generalization ability. Furthermore, SFCN is slightly more outstanding than FCN and LSTM.

Although SFCN is slightly inferior to the machine learning models on the latency, it is still tolerant in our scenario. Considering the outstanding performance on recognition accuracy and generalization ability comparing with other DL models, we finally choose SFCN as the core classifier of the system.

4.2 Determination of sampling frequency

After the model choosing step, the whole system was implemented and tested on the Android platform. Since the sampling frequency of WISDM is fixed, here we utilize SCD dataset. We collect activity data of different frequencies through multi-sampling and sub-sampling. We extracted the statistical features shown in Table 2 from the data blocks of all 8 subjects as train and test data to explore the relation between accuracy and sampling frequency. The experiment utilized 10-fold cross validation method to obtain the final results as shown in Table 6.

Table 6

Accuracy vs Sampling frequency of different activities

Activity	10Hz	20Hz	25Hz	30Hz	40Hz	50Hz
Sitting	1.00	1.00	1.00	1.00	1.00	1.00
Standing	0.89	0.91	0.93	0.94	0.93	0.94
Lying	1.00	1.00	1.00	1.00	1.00	1.00
Walking	0.77	0.89	0.89	0.90	0.92	0.93
Upstairs	0.66	0.78	0.82	0.87	0.93	0.96

Downstairs	0.61	0.69	0.75	0.78	0.81	0.84
Running	0.83	0.86	0.87	0.85	0.87	0.89
Quick Walk	0.71	0.78	0.83	0.85	0.92	0.98
Total	0.81	0.86	0.89	0.90	0.92	0.94

It is obvious that when sampling frequency down to 10Hz, the recognition accuracy of static activities remains almost 100%, while the negative effect on dynamic activity is significant. The minimum sampling frequency is determined by real scene. Considering that both the total and dynamic activity recognition accuracy achieve a remarkable decline (by 5% on average) when sampling frequency drops from 20Hz to 10Hz, we set the minimum sampling frequency as 20Hz, which means if the activity changes little, the sampling frequency will be reduced from 50Hz to 20Hz to save energy while the accuracy remained over 85%.

4.3 Settings of pre-classification strategies

In this part, we determine the parameters of the pre-classification methods through experiments.

First, we try to find a proper value of β . The parameter conf affects the HAR accuracy and also the call frequency of SFCN. A proper value of conf can make full use of pre-classification results and effectively reduce the high delay and energy consumption caused by SFCN. We set different value of β , which directly relates to conf , to obtain the change of accuracy and the call frequency of SFCN model (shown in Fig. 6). The experiment is repeated 3 times to exclude the influence of random factors, and the mean value is taken as final results. According to the results, when $\beta = 90$ and sampling frequency = 50Hz, the accuracy was over 90% while the call frequency of SFCN model was reduced to 10%. When $\beta = 50$ and sampling frequency = 20Hz, the accuracy still meets requirement and also reduce the call frequency of SFCN model by half. Therefore, in the system, the parameter set ($\beta = 90$, SF=50Hz) and ($\beta = 50$, SF=20Hz) are applied to online recognition stage.

Figure 6 Accuracy vs Call frequency of DL on different β

Second, we have discuss the influence of λ and δ on the pre-classification strategy, and the determination process of the two parameters is given in this part. We use the Android app mentioned above to collect daily data and recognize activities of 8 subjects. When the app

ran, the analysis tool PowerTutor is employed to record energy consumed. We also record the time delay of a recognition task by Android API.

We utilized grid search for parameter tuning. There is no significant change in accuracy or energy consumption when $\lambda < 10$ or $\lambda > 50$, so the value range is set as $\{10, 20, 30, 40, 50\}$. Similarly, the δ is set as $\{0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. The results are shown in Fig. 7.

(a)

(b)

Figure 7 Influence of λ and δ on Accuracy(a) and Power Consumption(b)

It is clear that when δ changed from 0.5 to 0.4, power consumption reduction plummeted from 39.6% to 35.4%. According to the experiment result, the δ needs to remain at 0.5 to keep an acceptable energy-efficient effect while the recognition accuracy is still above 84%. Meanwhile, when λ changes from 10 to 20, the average power consumption reduction decreases by 13.4% due to larger λ will introduce more time window for computing and so to increase the value of σ_t and also the power consumption indirectly. Based on this, we finally set $(\delta = 0.5, \lambda = 10)$.

4.4 Overall system performance

In this section, we conduct experiments to show the overall system performance. The DL in this section is just the SFCN model we propose in the above parts.

We first test the PC-DL model on the PC platform using the SCD dataset. Comparing with DL, the overall accuracy of PC-DL is lower. Still, it remains an 84% overall accuracy, which is a better performance than the traditional learning model. Meanwhile, the time delay of PC-DL

reduces to about 0.6ms per sample, which is a large promotion comparing with DL model. We finally perfect the Android app and use the parameter set for testing. We take 2 new subjects, who have similar physical features with the former 8 ones, for data collection and activity recognition. Each activity lasts for 1 minute and repeats for 10 times. A single experiment contained two parts, one for pure DL model and another introduces the proposed pre-classification strategy with DL model (PC-DL) to recognize activities. Energy consumption, time delay and recognition accuracy are recorded during the experiment. The results are shown in Table 7.

Table 7

Overall system performance

Activity	Power consumption/J			Time delay/ms			Recognition Accuracy		
	DL	PC-DL	Reduction	DL	PC-DL	Reduction	DL	PC-DL	Reduction
Sitting	31.5	19.9	0.37	1299.3	773.0	0.41	1.000	1.000	0.000
Standing	31.6	13.2	0.58	1274.6	453.9	0.64	0.947	0.920	0.027
Lying	28.3	8.2	0.71	1261.4	265.5	0.79	1.000	1.000	0.000
Walking	31.2	24.9	0.20	1297.8	832.1	0.36	0.863	0.817	0.046
Up stairs	27.9	13.1	0.53	1260.0	514.7	0.59	0.785	0.756	0.029
Down stairs	33.4	4.3	0.87	1291.8	139.2	0.89	0.777	0.730	0.047
Running	30.3	36.2	-0.19	1313.3	1336.4	-0.02	1.000	0.973	0.027
Quick Walk	38.5	8.3	0.78	1307.0	290.0	0.78	0.886	0.826	-0.066
Average	31.6	16.0	0.49	1288.1	575.6	0.55	0.907	0.878	0.014

According to the records, the energy consumption of PC-DL decreases by 49% on average due to it can reduce the computing load and the sampling frequency of mobile device. In addition, the PC-DL reduced the time delay by 55% on average. Meanwhile, the recognition accuracy is reduced but still remained at 87.78%, the transition diagram helps slightly reducing the misclassification ,because the errors always happens between linked activities in the diagram.

Moreover, to further verify the advantage of PC-DL on energy-efficiency we test the battery life of two Android phones in 3 scenarios: 1) Recognition with DL. 2) Recognition with PC-DL. 3) No recognition. To ensure the initial conditions are identical, we set the phone power 100% before start the experiment. We require users to avoid other operation (such as watching video or playing game) on the phone during the experiment. Figure 8 shows the energy consumption of the two smartphones in 3 scenarios. According to the figure, comparing with DL model, the PC-DL method is significantly reduced the power by 31%-39%.

(a)

(b)

Figure 8 Battery life of two users in 3 scenarios

5 Conclusion and Future work

In this work, we develop a lightweight HAR system for recognizing simple daily activities based on PC-DL model. We utilize convolutional network as the core classification algorithm. Meanwhile, we introduce the pre-classification strategies to reduce energy consumption and time delay by lower the algorithm complexity and sampling frequency of the sensors in smart phones.

In terms of recognition algorithm design, we compress and simplify the AlexNet for HAR task so as to ensure the over performance of the system. In addition, a transition diagram and the pre-classification strategy based on clustering-center are proposed. The clustering-center calculates the Euclidean distance between the objective activity and the whole set to achieve pre-classified results. The strategy avoids the classification mode on statistics and greatly reduces the hardware resource occupation and time complexity of the algorithm. On the other hand, we also propose the activity change based on activity intensity, which reflects the

probability of maintaining the current activity state. The system changes sampling frequency in real-time if the activity change increases over the threshold. Considering that HAR is a long duration process, the control strategy allows the system to operate at a lower sampling frequency in most of the time, which can reduce both the energy consumption and computation load of the mobile phone.

By deploying the HAR app developed on the Android platform, this research verifies the good performance of the PC-DL algorithm on energy-efficiency and time delay: the overall accuracy of recognizing 8 activities remains about 85%, while energy consumption is reduced by 49% and the time delay is reduced by 55%. In the long-time test, the experiment showed that the PC-DL algorithm can reduce the consumption of battery by 31%-39% compared with the DL algorithm.

In future work, we will expand the activity dataset. Also, we will consider multiple kinds of sensors for complex activity recognition.

Reference:

- [1] Cornacchia, Maria, et al. "A Survey on Activity Detection and Classification Using Wearable Sensors." *IEEE Sensors Journal* 17.2(2016):386-403. DOI: 10.1109/JSEN.2016.2628346
- [2] Chen, Lien Wu, et al. "Intelligent file transfer for smart handheld devices based on mobile cloud computing." *International Journal of Communication Systems* (2015). DOI: 10.1002/dac.2947.
- [3] Altun, Kerem, and B. Barshan. "Human Activity Recognition Using Inertial/Magnetic Sensor Units." *International Conference on Human Behavior Understanding Springer-Verlag*, 2010:38-51. DOI: 10.1007/978-3-642-14715-9_5
- [4] Vermeulen, J.;Willard, S.; Aguiar, B.; DeWitte, L.P. Validity of a Smartphone-Based Fall Detection Application on Different Phones Worn on a Belt or in a Trouser Pocket. *Assist. Technol.* 2015, 27, 18. DOI: 10.1080/10400435.2014.949015
- [5] Chernbumroong, Saisakul, C. Shuang, and H. Yu. "A practical multi-sensor activity recognition system for home-based care." *Decision Support Systems* 66.C(2014):61-70. DOI: 10.1016/j.dss.2014.06.005
- [6] Wang, Rui , et al. "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones." *Acm International Joint Conference on Pervasive & Ubiquitous Computing ACM*, 2014. DOI: 10.1145/2632048.2632054
- [7] Oshin, T.O.; Poslad, S.; Zhang, Z. Energy-Efficient Real-Time Human Mobility State Classification Using Smartphones. *IEEE Trans. Comput.* 2015, 64, 1680–1693 DOI: 10.1109/TC.2014.2339846
- [8] Morillo, L.M.S.; Gonzalez-Abril, L.; Ramirez, J.A.O. Low Energy Physical Activity Recognition System on Smartphones. *Sensors* 2015, 15, 5163–5196 DOI: 10.3390/s150305163
- [9] Fullerton Elliott, B. Heller, and M. Munoz-Organero. "Recognising human activity in free-living using multiple body-worn accelerometers." *IEEE Sensors Journal* PP.99(2017):1-1.DOI: 10.1109/jsen.2017.2722105
- [10] Thanh, P Noi , and M. Kappas . "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. " *Sensors* 18.1(2018):18.
- [11] Wang, Jidong, et al. "A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree." *Applied Sciences* 8.5(2018):689-.
- [12] Zhang, S., et al. "Efficient kNN Classification With Different Numbers of Nearest Neighbors. " *IEEE Transactions on Neural Networks & Learning Systems* 29.5(2018):1774-1785.
- [13] Jeong, Chi Yoon , and M. Kim . "An Energy-Efficient Method for Human Activity Recognition with Segment-Level Change Detection and Deep Learning." *Sensors* 19.17(2019):3688.

- [14] S. W. Pienaar and R. Malekian, "Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture," 2019 IEEE 2nd Wireless Africa Conference (WAC), Pretoria, South Africa, 2019, pp. 1-5, doi: 10.1109/AFRICA.2019.8843403.
- [15] Pansiot, Julien, et al. "Ambient and Wearable Sensor Fusion for Activity Recognition in Healthcare Monitoring Systems." International Workshop on Wearable and Implantable Body Sensor Networks, Bsn 2007, March 26-28, 2007, Rwth Aachen University, Germany DBLP, 2007:208-212. DOI: 10.1007/978-3-540-70994-7_36
- [16] Minnen, David , et al. "Recognizing Soldier Activities in the Field." International Workshop on Wearable & Implantable Body Sensor Networks DBLP, 2008. DOI: 10.1007/978-3-540-70994-7_40
- [17] Angelini, Leonardo, et al. "Designing a desirable smart bracelet for older adults." ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication ACM, 2013:425-434. DOI: 10.1145/2494091.2495974
- [18] Akhavian, Reza, and A. H. Behzadan. "Smartphone-based construction workers' activity recognition and classification." Automation in Construction 71(2016):198-209. DOI: 10.1016/j.autcon.2016.08.015
- [19] Tran, Duc Ngoc , and D. D. Phan . "Human Activities Recognition in Android Smartphone Using Support Vector Machine." International Conference on Intelligent SystemsIEEE, 2017. DOI: 10.1109/ISMS.2016.51
- [20] Bisio, Igor , et al. "Smartphone-based user Activity Recognition Method for Health Remote Monitoring Applications. " Mobile & Pervasive Computing, Peccs 2012.
- [21] Xie, Yutong, et al. "Fusing Texture, Shape and Deep Model-Learned Information at Decision Level for Automated Classification of Lung Nodules on Chest CT." Information Fusion 42(2017). <https://doi.org/10.1016/j.inffus.2017.10.005>
- [22] Galea, Christian, and R. A. Farrugia. "Matching Software-Generated Sketches to Face Photos with a Very Deep CNN, Morphed Faces, and Transfer Learning." IEEE Transactions on Information Forensics & Security PP.99(2017):1-1. DOI: 10.1109/TIFS.2017.2788002
- [23] Chen, M. C., et al. "Deep Learning to Classify Radiology Free-Text Reports." Radiology (2017):171115. DOI: 10.1148/radiol.2017171115
- [24] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "ImageNet classification with deep convolutional neural networks." International Conference on Neural Information Processing Systems Curran Associates Inc. 2012:1097-1105. DOI: 10.1145/3065386
- [25] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." (2015):770-778. DOI: 10.1109/CVPR.2016.90
- [26] Yang, Jian Bo, et al. "Deep convolutional neural networks on multichannel time series for human activity recognition." International Conference on Artificial Intelligence AAAI Press, 2015:3995-4001.
- [27] Hammerla, Nils Y., S. Halloran, and T. Ploetz. "Deep, Convolutional, and Recurrent Models for Human Activity Recognition using Wearables." Journal of Scientific Computing 61.2(2016):454-476. DOI: 10.1007/s10915-014-9833-1
- [28] Plötz, T, N. Y. Hammerla , and P. Olivier . "Feature learning for activity recognition in ubiquitous computing." International Joint Conference on Ijcai DBLP, 2011. DOI: 10.5591/978-1-57735-516-8/IJCAI11-290
- [29] T. Zebin, M. Sperrin, N. Peek and A. J. Casson, "Human activity recognition from inertial sensor time-series using batch normalized deep LSTM recurrent networks," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI, 2018, pp. 1-4, doi: 10.1109/EMBC.2018.8513115.
- [30] Wang, Lu Kun , and R. Y. Liu . "Human Activity Recognition Based on Wearable Sensor Using Hierarchical

Deep LSTM Networks." *Circuits Systems and Signal Processing* (2019).

- [31] Reddy, Sasank , et al. "Using mobile phones to determine transportation modes." *ACM Transactions on Sensor Networks* 6.2(2010):1-27. DOI: 10.1145/1689239.1689243
- [32] Morillo, Luis , et al. "Low Energy Physical Activity Recognition System on Smartphones." *Sensors* 15.3(2015):5163-5196. DOI: 10.3390/s150305163
- [33] Yan, Zhixian , et al. "Energy-Efficient Continuous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach." *International Symposium on Wearable Computers IEEE Computer Society*, 2012. DOI: 10.1109/ISWC.2012.23
- [34] Qi X, Li Y, Keally M, et al. AdaSense: Adapting sampling rates for activity recognition in Body Sensor Networks 2013:163-172. DOI: 10.1109/RTAS.2013.6531089
- [35] Rault, Tifenn , et al. "A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications." *Pervasive & Mobile Computing* 37.C(2017):23-44. DOI: 10.1016/j.pmcj.2016.08.003
- [36] Chu, David , et al. " [ACM Press the 9th ACM Conference - Seattle, Washington (2011.11.01-2011.11.04)] Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems - SenSys '11 - Balancing energy, latency and accuracy for mobile sensor data classification." (2011):54.
- [37] Gordon, Dawud , et al. "Energy-Efficient Activity Recognition Using Prediction." 2012 16th International Symposium on Wearable Computers IEEE, 2012.
- [38] Kwapisz, Jennifer R. , G. M. Weiss , and S. A. Moore . "Activity recognition using cell phone accelerometers." *Acm Sigkdd Explorations Newsletter* 12.2(2011):74-82.
- [39] Abedin, Alireza & Ranasinghe, Damith & Rezatofighi, Hamid & Shi, Qinfeng. (2019). *SparseSense: Human Activity Recognition from Highly Sparse Sensor Data-streams Using Set-based Neural Networks*.
- [40] Ainsworth, B. E. , et al. "Compendium of Physical Activities: an Update of Activity Codes and MET Intensities." *Medicine & Science in Sports & Exercise* 32.9 Suppl(2000):498-504. DOI: 10.1249/00005768-199301000-00011
- [41] Jeong, Chi Yoon , and M. Kim . "An Energy-Efficient Method for Human Activity Recognition with Segment-Level Change Detection and Deep Learning." *Sensors* 19.17(2019):3688.

Affiliations

Shi Junhao: Doctoral candidate.

Email: shijunhao@hit.edu.cn

Zuo Decheng: Professor, doctoral tutor. Assistant dean of the institute of computer science and technology and director of the fault-tolerant and mobile computing research center of HIT.

Email: zuode@hit.edu.cn

Zhang Zhan: Associate professor, master tutor. Researcher of the fault-tolerant and mobile computing research center of HIT.

Email: zz@ftcl.hit.edu.cn

Figures

Figure 1

System architecture

Figure 1

System architecture

Figure 2

Process of convolution

Figure 2

Process of convolution

(a)

(b)

Figure 3

Relation between DL model recognition accuracy and training epochs (a) and kernel size (b)

(a)

(b)

Figure 3

Relation between DL model recognition accuracy and training epochs (a) and kernel size (b)

Figure 4

Confusion Matrix of 7 classifiers (WISDM)

Figure 4

Confusion Matrix of 7 classifiers (WISDM)

Figure 5

Time delay of the classifiers

Figure 5

Time delay of the classifiers

Figure 6

Accuracy vs Call frequency of DL on different β

Figure 6

Accuracy vs Call frequency of DL on different β

(a)

(b)

Figure 7

Influence of λ and δ on Accuracy(a) and Power Consumption(b)

(a)

(b)

Figure 7

Influence of λ and δ on Accuracy(a) and Power Consumption(b)

(a)

(b)

Figure 8

Battery life of two users in 3 scenarios

(a)

(b)

Figure 8

Battery life of two users in 3 scenarios