International Journal of Smart Home
Vol. 10, No. 11, (2016), pp. 95-108
http://dx.doi.org/10.14257/ijsh.2016.10.11.09

# Uncertainty Sampling Based Posterior Probability Extreme Learning Machine for Human Activity Recognition

Xiaopeng Niu[1*], Zhiliang Wang[1], Wendong Xiao[2] and Zhigeng Pan[3]

[1]*School of Computer and Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[2]*School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China*
[3]*Digital Media and HCI Center, Hangzhou Normal University, Hangzhou, 311121, China*
[*]*supernxpeng@126.com*

### *Abstract*

*Human activity recognition is a main research area of context-aware computing, and is widely used in many applications, such as smart home and elderly care. Smartphone-based human activity recognition is very popular by making use of the embedded inertial sensors. However, there exists the problems of misclassification activities, and how to effectively apply the model trained by known users to new users. To solve these two problems, in this paper, we proposed a novel approach, Uncertainty Sampling based posterior Probability Extreme Learning Machine (USP-ELM), by introducing two strategies: first, we transfer the actual outputs of ELM to posterior probabilities for each instances, and then use uncertainty sampling strategy for confidence level assignment to adapt the training model and improve the classification accuracy. Experimental results show that the proposed approach is more efficient, compared with the existing ELMs.*

***Keywords:*** *Human Activity Recognition, Context-awareness, Smartphone, ELM, Posteriori Probability, Uncertainty Sampling*

## 1. Introduction

Human activity recognition is a main research field of Context-aware computing. Sensing and recognizing human activities (walking, lying, standing, sitting, eating, etc.) are very important in many context-aware applications, such as smart home, eldercare, and healthcare [1-3]. There are three approaches to research human activity recognition: smartphone based, wearable sensors based and development board based. In the era of the Internet of Thing, due to the evolution of smartphones with a variety of sensors, the smartphone-based approach has become very popular [7].

Human activities are categorized in three main groups: short events, basic activities， and complex activities. Short events are comprised of Postural Transitions (such as sit-to-stand). In many existing research approaches, transitions between activities were not considered, and this can affect the performance of recognition system. There are two common misclassification types, one occurs during basic activities (such as sitting and standing), the other occurs during postural transitions [5]. How to effectively apply training model on known user activities to recognize unknown user activities is also a problem when research human activity recognition.

In this paper, we proposed an approach, USP-ELM to solve the above problems (misclassification, recognize unknown new user activities) using the public dataset [5-6]. First, we used the training data to train an initial ELM model. Second, we used the testing

---

[*]Corresponding Author

ISSN: 1975-4094 IJSH

data to test the initial model, and the actual outputs of the model were transferred to posterior probabilities of each testing data. Third, we used Least Confidence and Margin sampling strategies [8] to adapt the initial model in order to improve the accuracy of the recognition system. Experiment results showed that the proposed approach was very efficient.

The rest of this paper is organized as follows. Section 2, reviews the related work and preliminary. Section 3, presents the proposed approach. Section 4 presents the experiment results. Section 5 draws the conclusion and future work.

## 2. Related Work and Preliminary

### 2.1. Related Work

Nowadays, there are a lot of researchers are working on Human Activity Recognition, usually using smartphone or wearable sensors.

In order to solve the problem of device displacement, Chen *et. al.,* [2] proposed a fast, robust activity recognition model. They used a wearable device to collect acceleration data, used Principal Component Analysis (PCA) as a feature selection approach, and used ELM to train an initial model. In the testing state, they used a confidence level strategy to retrain and adapt the initial model. The experiment results showed that this strategy can effectively solve the problem of device displacement. Deng *et. al.,* [3] proposed a TransRKELM (Transfer learning Reduced Kernel Extreme Learning Machine) model to solve the cross-person problem. They used Reduced Kernel Extreme Learning Machine as a classifier model, and adopt Chen *et. al.,* [2] confidence level strategy to solve the cross-person problem. Reyes-Ortiz *et. al.,* [5] proposed a Transition-Aware Human Activity Recognition (TAHAR) system architecture for the recognition of human activities using smartphones, they used posterior probability SVM and filtering strategies to solve the problem of postural transitions between static activities in order to improve the recognition performance.

Xiao *et. al.,* [9] proposed an activity recognition model based on Kernel Discriminant Analysis (KDA) and ELM, they used KDA as a feature selection approach, and ELM as a classifier model. He *et. al.,* [10] proposed an activity recognition model based on Generalized Discriminant Analysis (GDA) and Relevance Vector Machine (RVM), they used GDA as a feature selection approach, and RVM as a classifier model.

### 2.2. Extreme Learning Machine

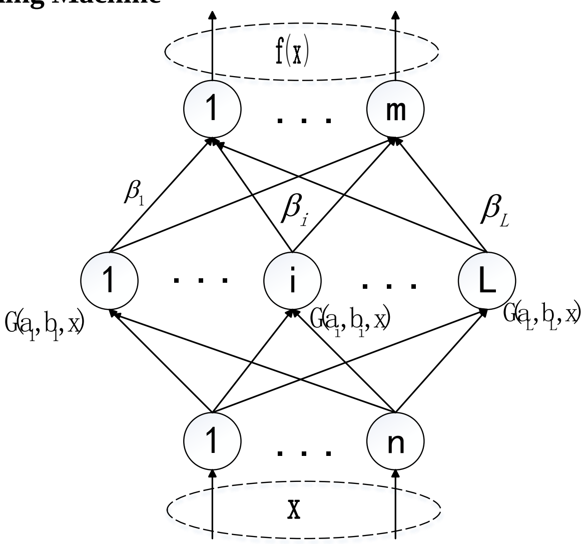


**Figure 1. The Structure of ELM**

Extreme Learning Machine (ELM) proposed by Huang *et. al.*, [14] is a single-hidden layer feedforward neural network (SLFN) with random input weights and random hidden nodes. The hidden layer of ELM need not be tuned, and the input weights and the hidden nodes are randomly chosen. ELM doesn't need to iteratively tune the parameters between the input layer and the hidden layer, but denotes their values randomly, then only calculates the weights connecting the hidden layer and the output layer by least-square method. So ELM algorithm has the properties of fast learning speed, light computational costs, high accuracy, and good generalization performance [9, 13-17].

The structure of ELM is shown as Figure 1. For N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\boldsymbol{x}_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in \boldsymbol{R}^n$ and $\boldsymbol{t}_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in \boldsymbol{R}^m$, the activity function of hidden layer neurons is $g(x)$, and there are L hidden layer nodes, $\boldsymbol{a}_i = [a_{i1}, a_{i2}, ... a_{in}]^T$ is the weight vector connecting the $i$ th hidden node and the input nodes, $\beta_i = [\beta_{i1}, \beta_{i2}, ... \beta_{im}]^T$ is the weight vector connecting the $i$ th hidden node and the output nodes, $b_i$ is the threshold of $i$ th hidden node. $G(\boldsymbol{a}_i, \mathbf{b}_i, \mathbf{x})$ is the output of the $i$ th hidden node and is calculated by Equation (1), and the output of ELM is calculated by Equation (2).

$$G(\boldsymbol{a}_i, \mathbf{b}_i, \mathbf{x}) = g(\mathbf{a}_i \cdot \mathbf{x} + b_i), \mathbf{a}_i \in \mathbf{R}^n, b_i \in \boldsymbol{R} \tag{1}$$

$$\boldsymbol{f}(x_j) = \sum_{i=1}^{L} \beta_i G(\mathbf{a}_i, b_i, x_j) = t_j, \beta_i \in \boldsymbol{R}^m, j = 1, 2, ..., N \tag{2}$$

Equation (2) can be written compactly as Equation (3)

$$\boldsymbol{H}\beta = \boldsymbol{T} \tag{3}$$

where

$$\begin{aligned} &\mathbf{H}(\mathbf{a}_1, \dots, \mathbf{a}_L, b_1, \dots, b_L, \mathbf{x}_1, \dots, \mathbf{x}_N) \\ &= \begin{bmatrix} g(\mathbf{a}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{a}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(\mathbf{a}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{a}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} = [\mathbf{h}(\mathbf{x}_1), ..., \mathbf{h}(\mathbf{x}_N)]^T_{L \times N} \end{aligned} \tag{4}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m} \text{ and } \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \tag{5}$$

$\mathbf{H}$ is called the hidden layer output matrix of the neural network, the $i$ th column of $\mathbf{H}$ is the $i$ th hidden node output with respect to inputs $\mathbf{x}_1, \dots, \mathbf{x}_N$ [14]. The output weights can be calculated by Equation (6)

$$\beta = \mathbf{H}^{\dagger}\mathbf{T} \tag{6}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of matrix $\mathbf{H}$.

The output function of ELM is [12-15]

$$\mathbf{f}(\mathbf{x}) = \begin{cases} \mathbf{h}(\mathbf{x})\mathbf{H}^T\left(\dfrac{\mathbf{I}}{C} + \mathbf{H}\mathbf{H}^T\right)^{-1}\mathbf{T}, \text{ when } N < L \\ \mathbf{h}(\mathbf{x})\left(\dfrac{\mathbf{I}}{C} + \mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{T}, \text{ when } N \geq L \end{cases} \tag{7}$$

$f_j(\mathbf{x})$ is the output function of the $j$ th output node, $\mathbf{f}(\mathbf{x}) = \left[f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})\right]^T$, the predicted class label of sample $\mathbf{x}$ is

$$label(\mathbf{x}) = \arg \max_{i \in (1,\ldots,m)} f_i(\mathbf{x}) \tag{8}$$

## 3. Proposed Approach

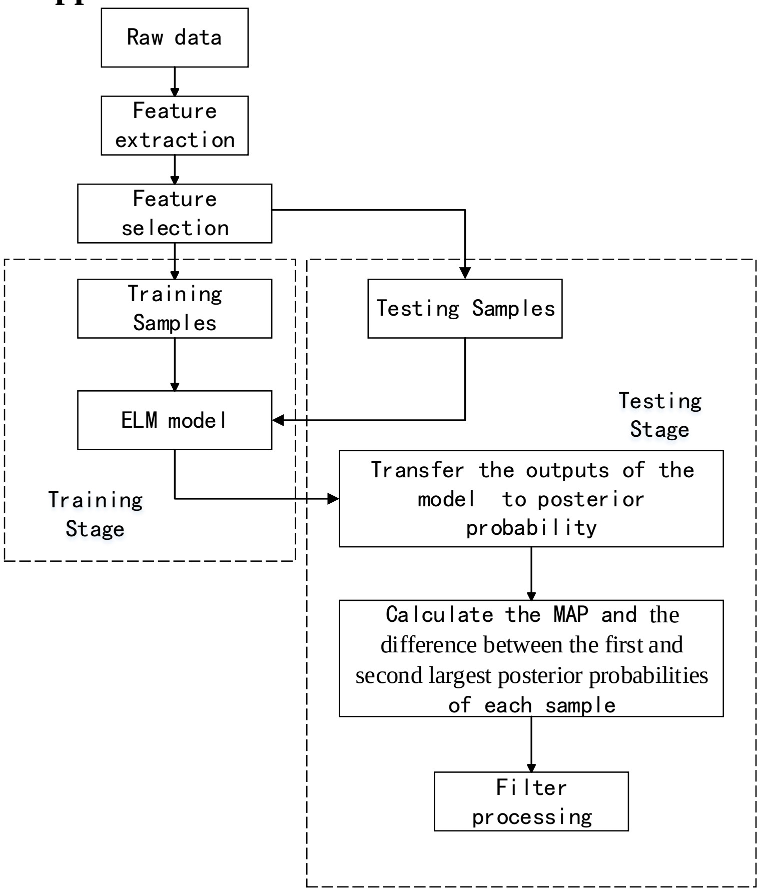


**Figure 2. The Architecture of the Proposed Approach**

Figure 2, shows the architecture of the proposed approach. The proposed approach combined posterior probability ELM and Uncertainty Sampling strategy [8], in order to solve the problem of common misclassification, and deal with the problem of unknown user activity recognition.

First, features were extracted from raw data which were collected from accelerometer and gyroscope by smartphone, and the subset of features was selected in order to reduce the dimension of features, and then it was divided into two parts (training samples, testing samples).

 

Second, the training samples were used to train an initial ELM model, then the trained model was tested by the testing samples, and the actual outputs were transferred to posterior probabilities.

Third, the strategy of least confidence [8] was used to calculate the difference between the first and second largest posterior probabilities of testing samples, and the strategy of Margin Sampling [8] was used to calculate the Maximum A Posterior Probability (MAP) [4]. The smaller the difference is, the more uncertainty the corresponding instance is [12], and the smaller the MAP is, the more uncertainty the corresponding instance is.

In the filter processing stage, the MAP and the difference between the first and second largest posterior probabilities of each testing sample were combined as sample's confidence level, and the threshold of the confidence level was set. When sample's confidence level is less than threshold, the predicted class label is the index which has the second largest posterior probability of the sample. The samples whose confidence levels were larger than threshold will be added into training samples to retrain the initial ELM model.

### 3.1. Posterior Probability Extreme Learning Machine

Ruck *et al.*, [19] proposed that the outputs of the multilayer perceptron are approximating the a posteriori probability functions of the classes being trained. Platt [20] proposed posteriori probability Support Vector Machines (SVMs), which can transfer the actual output $f(x)$ into posteriori probabilities of the instance $x$ by Equation (9), and the parameters A and B are fit using maximum likelihood estimation from training set.

$$P(y = 1|f(x)) = \frac{1}{1 + \exp(Af(x) + B)} \tag{9}$$

For a given number of class m and a test sample $\mathbf{x}$, the he predicted class label of sample $\mathbf{x}$ is

$$label(\mathbf{x}) = \arg\max_{i \in (1,\dots,m)} p_i(\mathbf{x}) \tag{10}$$

where $p_i(\mathbf{x})$ is posteriori probabilities outputs of each SVM [5].

Compare Equation (8) and Equation (10), we can find that there are some links between ELM and posteriori probability SVM. Yu *et al.*, [12-13] adopt Platt's approach, proposed posteriori probability ELM by Equation (11)

$$P(y = 1|f_i(x)) = \frac{1}{1 + \exp(-f_i(x))} \tag{11}$$

where $f_i(x)$ denotes the actual output of the $i$ th output node for the sample $x$.

### 3.2. Uncertainty Sampling

Perhaps Uncertainty Sampling is the simplest and most commonly used approach in active learning, this approach uses probabilistic learning model, and it has three strategies: Least Confidence, Margin Sampling, and Entropy [8]. In this paper, the strategies of Least Confidence and Margin Sampling are used in the proposed approach.

1) Least Confidence

For multi-class labels, the Least Confidence is obtained by Equation (12)

$$x_{LC} = \arg\max_{x} 1 - \mathrm{P}_\theta(y|x) \tag{12}$$

$$y = \arg\max_{y} \mathrm{P}_{\theta}(y|x) \tag{13}$$

$y$ is the class label with the highest posterior probability under the model $\theta$. The shortcoming of the least confidence strategy is that this strategy only considers information about the most probable labels, throw away information about the remaining label distribution [8].

2) Margin Sampling

Margin Sampling is obtained by Equation (14)

$$x_{MS} = \arg\min_{x} \mathrm{P}_{\theta}(y_1|x) - \mathrm{P}_{\theta}(y_2|x) \tag{14}$$

where $y_1$ and $y_2$ are the first and second most probable class labels under the model, respectively. The margin between the first and second most probable class is calculated by Equation (15) [16]

$$x_{M} = \mathrm{P}_{\theta}(y_1|x) - \mathrm{P}_{\theta}(y_2|x) \tag{15}$$

Margin Sampling solves the shortcoming of the least confidence by using the second largest posteriori probability of class labels.

## 4. Experimental Results

### 4.1. Experimental Dataset and Feature Selection

In this paper, we used the dataset which is collected from a group of 30 volunteers within an age bracket of 19-48 years, and it consists of 10929 instances [5]. They performed a protocol of activities composed of six basic activities using a smartphone which has a 3-axis accelerometer and a gyroscope: three static postures (standing, sitting, lying) and three dynamic activities (walking, walking downstairs and walking upstairs). The experiment also included postural transitions that occurred between the static postures (stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, and lie-to-stand).

The sensor signals (accelerometer and gyroscope) were sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). From each window, a vector of 561 features was extracted by calculating variables from the time and frequency domain. The dataset was randomly partitioned into two sets, where 70% selected as training set and 30% as testing set. The dataset can be downloaded in [21].

The dataset has 12 activity classes: 6 basic activity classes and 6 postural transition classes. We adopt the approach proposed by Reyes-Ortiz *et al.,* [5] that considered 6 postural transition classes as 1 class which is called postural transitions. So, in this way, there are 7 activity classes.

The feature set has 561 features including 272 time-domain features and 289 frequency-domain features. Reyes-Ortiz [4] proposed that the frequency-domain features do not largely affect recognition performance when compared with time-domain features. We used ELM as classifier, and the recognition performance is shown in Table 1. So, we selected 272 time-domain features as feature set.

**Table 1. The Recognition Performance with Different Feature Sets**

| Feature set | Number of features | Testing accuracy |
|---|---|---|
| time-domain | 272 | 95.5% |
| total | 561 | 95.3% |

 

### 4.2. Improve the Recognition Performance with Posterior Probability ELM

In order to validate the effectiveness of the proposed approach, we used the original ELM model as a baseline. Through experiments, we set 1800 hidden layer nodes, and the regularized parameter C is $2^{-5}$, the confusion matrix of testing set is shown as Table 2.

**Table 2. Confusion Matrix of Testing Set Using the Original ELM**

| | WK | WU | WD | SI | ST | LY | PT |
|---|---|---|---|---|---|---|---|
| WK | 488 | 4 | 4 | 0 | 0 | 0 | 0 |
| WU | 24 | 445 | 2 | 0 | 0 | 0 | 0 |
| WD | 4 | 12 | 404 | 0 | 0 | 0 | 0 |
| SI | 0 | 3 | 0 | 455 | 49 | 0 | 1 |
| ST | 0 | 1 | 0 | 28 | 526 | 0 | 1 |
| LY | 0 | 0 | 0 | 0 | 5 | 540 | 0 |
| PT | 2 | 2 | 0 | 2 | 1 | 0 | 159 |

WK: Walking, WU: Walking-upstairs, WD: Walking-downstairs, SI: Siting, ST: Standing, LD: Laying, PT: Postural Transitions

From Table 2, we can find that there are two types of misclassification during recognition: one type occurs during basic activities (WK, WU, SI, and ST), and one type occurs between Postural Transitions and basic activities.

From Equation (8), we can find that the predicted class label of original ELM is the index which has the maximum output among all outputs of each instance, and the other instances are not considered. But, human activities can be considered as a sequence of correlated event [4], only use the information of one instance to predict class label is not enough. So, the proposed approach used Equation (11) to transfer the original ELM to posterior probability ELM.

As the inputs are a sequence of activity instances, so the outputs of posterior probability ELM can be considered as an activity matrix which includes posterior probabilities from neighboring instances. The proposed approach used the MAP and the value of Equation (15) to measure the uncertainty of each instance.

Figure 3, shows an example of the largest and the second largest posterior probability outputs of instances. One misclassification occurs between 10 and 20s (between Postural Transition and WU), and one misclassification occurs between 70 and 80s (between ST and SI).

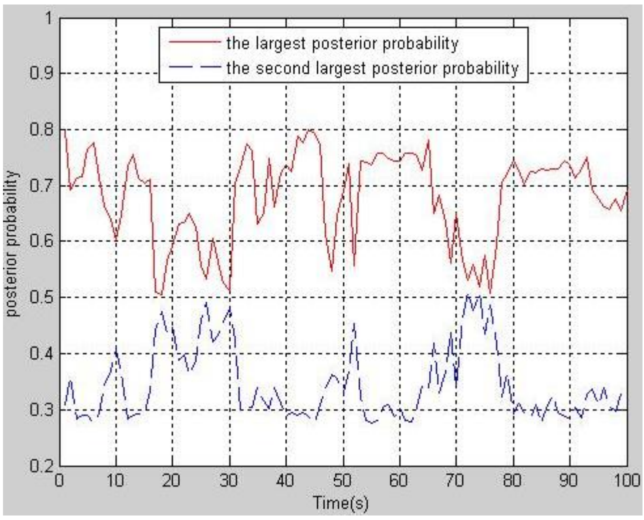


**Figure 3. An Example of the Largest and the Second Largest Posterior Probability Outputs of Instances**

 

From Table 2, we can find that two group basic activities that are the most likely misclassified: one group is WK and WU, one group is SI and ST. We set $\alpha_W$ , $\beta_W$ are the thresholds of WK, $\alpha_S$ , $\beta_S$ are the thresholds of ST. Take the group of SI and ST as an example, the proposed approach contains two steps:

Step1: for each instance, calculate the MAP and the value of Equation (15).

Step2: if the MAP is less than $\alpha_S$ , and the value of Equation (15) is less than $\beta_S$ , then the predicted class label is the index which has the second largest posterior probability of each instance.

We experimented 100 trials, Figure 4, shows the recognition performance of original ELM and proposed approach, we set $\alpha_W$ =0.85, $\beta_W$ =0.035, $\alpha_S$ =0.5, and $\beta_S$ =0.1. Table 3 shows the comparison of original ELM and proposed approach.

From Figure 4, and Table 3, we can find that proposed approach has higher testing accuracy and less deviation than original ELM. The proposed approach used Uncertainty Sampling strategy to measure the uncertainty of each testing sample. The experimental result showed that the more uncertainty of the testing sample is, the more likely it will be misclassified.

**Table 3. The Comparison of Original ELM and Proposed Approach**

| | original ELM | proposed approach |
|---|---|---|
| Testing Accuracy(%) | 95.5±0.35 | 95.9±0.26 |

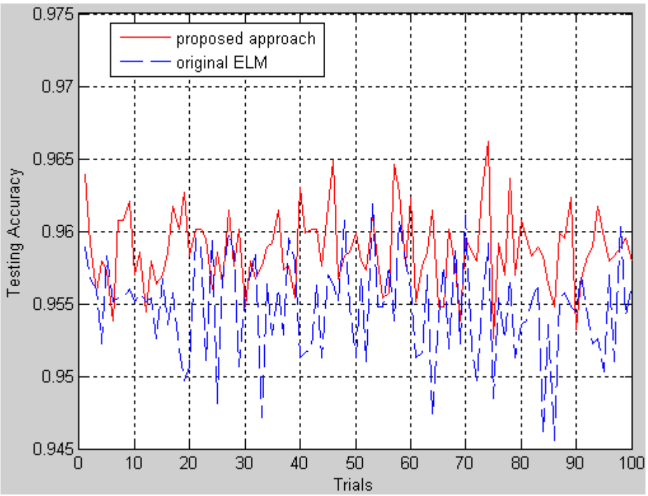


**Figure 4. The Recognition Performance of Original ELM and the Proposed Approach**

### 4.3. Improve the Recognition Performance with Posterior Probability ELM

How to effectively apply the model trained by known users to new users, is another problem when research human activity recognition. Chen *et. al.*, [2] and Deng *et. al.*, [3] proposed a model adaptation strategy by Equation (16) and Equation (17).

$$TY_i = TY_i - \min\left(TY_i\right) \tag{16}$$

$$confidence = \frac{\max\left(TY_i\right)}{\sum TY_i}, i = 1, 2, \ldots, m \tag{17}$$

 

where $TY_i$ is $i$ th output of ELM.

The confidence of Deng's [3] strategy approximate the normalized maximum output of each instance, the shortcoming of this strategy is only considers information about one class label, throw away information about the remaining class labels. In this paper, we proposed a more effective strategy by the MAP and Equation (15) using posterior probability ELM. We set $\lambda$, $\omega$ are the MAP and the value of Equation (15), respectively, and set $\theta$ is the threshold of Deng's strategy.

We adopt Deng's [3] experiment steps, and randomly selected 4 users which are denoted as user A, user B, user C, and user D from 30 users. In order to validate the effectiveness of the proposed approach, we divided these 4 users into three groups:

1) Group 1: user A, user B, and user C
2) Group 2: user A, user B, and user D
3) Group 3: user B, user C, and user D

Take group 1 as an example, the experiment steps are as follows [3].

The datasets of these users are $D_A$, $D_B$ and $D_C$, respectively. Each dataset is randomly divided into two parts (70%, 30%), which are represented as $D_{A1}$ and $D_{A2}$, $D_{B1}$ and $D_{B2}$, and $D_{C1}$ and $D_{C2}$. We first assume that user A and user B are known users, and user C is a new one. $Train_{AB}$ which equals $D_{A1} \cup D_{B1}$, is used to train an initial ELM model. $D_{C1}$ is used to adapt the initial model to a new one. $Test_{AB}$ which equals $D_{A2} \cup D_{B2}$, is used to test the two model's classification capability on the known users. $D_{C2}$ is used to test the two model's classification capability on the new user. For the initial model and each test data in $D_{C1}$, if $\lambda$ and $\omega$ are larger than threshold, respectively, it will be added into a new dataset $RT_{C1}$. Then, using the updated training dataset $Train_{AB} \cup RT_{C1}$, to retrain the initial ELM model.

We set 1000 hidden layer nodes, and the regularized parameter C is $2^{-2}$. Through 100 experiments, we found out the optimal threshold $\theta$ in Deng's approach is 0.6, and the optimal values of proposed approach are as follows $\lambda = 0.6, \omega = 0.05$. The recognition performance comparison of Deng's approach and proposed approach are shown in Table 4-9. We first assumed that user A and user B are known users, user C is a new user. Table 4 shows the performance of original ELM model, Deng's approach and our proposed approach on new user. We can see that the performance of Deng's approach and our proposed approach are both better than original ELM, but the accuracy of proposed approach is higher than Deng's approach. Table 5 shows the performance of original ELM model, Deng's approach and our proposed approach on known users, the performance are the same as original ELM.

When user A and user C are known users, and user B is a new user, the performance comparison of Deng's approach and proposed approach are shown in Table 6-7. When user B and user C are known users, and user A is a new user, the performance comparison of Deng's approach and proposed approach are shown in Table 8-9.

From Table 4-9, we can see that the recognition performance of proposed approach in Group 1 is better than Deng's approach.

In order to validate the generality of proposed approach, we tested the other two groups in the same way, and the experimental results are shown in Table10-21. We can see that the performance of proposed approach is more effective than Deng's approach.

**Table 4. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Train data | $Train_{AB}$ | $Train_{AB}$+$RT_{C1}$ | $Train_{AB}$+ $RT_{C1}$ |
| Test data | $D_{C2}$ | $D_{C2}$ | $D_{C2}$ |
| Accuracy(%) | 90.6 | 91.4 | 92.7 |

Note: user A and user B are known users, user C is a new user

**Table 5. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{AB}$ | $Train_{AB}$+ $RT_{C1}$ | TrainAB+ $RT_{C1}$ |
| Test data | $Test_{AB}$ | $Test_{AB}$ | $Test_{AB}$ |
| Accuracy(%) | 99.7 | 99.7 | 99.7 |

Note: user A and user B are known users, user C is a new user

**Table 6. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{AC}$ | $Train_{AC}$+ $RT_{B1}$ | $Train_{AC}$+ $RT_{B1}$ |
| Test data | $D_{B2}$ | $D_{B2}$ | $D_{B2}$ |
| Accuracy(%) | 67.8 | 68.6 | 73.5 |

Note: user A and user C are known users, user B is a new user

**Table 7. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{AC}$ | $Train_{AC}$+ $RT_{B1}$ | $Train_{AC}$+ $RT_{B1}$ |
| Test data | $Test_{AC}$ | $Test_{AC}$ | $Test_{AC}$ |
| Accuracy(%) | 99.4 | 99.4 | 99.4 |

Note: user A and user C are known users, user B is a new use

**Table 8. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BC}$ | $Train_{BC}$+ $RT_{A1}$ | $Train_{BC}$+ $RT_{A1}$ |
| Test data | $D_{A2}$ | $D_{A2}$ | $D_{A2}$ |
| Accuracy(%) | 83.8 | 85.3 | 86.2 |

Note: user B and user C are known users, user A is a new user

**Table 9. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BC}$ | $Train_{BC}$+ $RT_{A1}$ | $Train_{BC}$+ $RT_{A1}$ |
| Test data | $Test_{BC}$ | $Test_{BC}$ | $Test_{BC}$ |
| Accuracy(%) | 98 | 98.1 | 98.3 |

Note: user B and user C are known users, user A is a new user

**Table 10. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Train data | $Train_{AB}$ | $Train_{AB}$+$RT_{D1}$ | $Train_{AB}$+ $RT_{D1}$ |
| Test data | $D_{D2}$ | $D_{D2}$ | $D_{D2}$ |
| Accuracy(%) | 90.8 | 94.6 | 95.3 |

Note: user A and user C are known users, user B is a new user

**Table 11. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{AB}$ | $Train_{AB}$+ $RT_{D1}$ | $Train_{AB}$+ $RT_{D1}$ |
| Test data | $Test_{AB}$ | $Test_{AB}$ | $Test_{AB}$ |
| Accuracy(%) | 99.6 | 99.3 | 99.3 |

Note: user A and user C are known users, user B is a new user

**Table 12. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{AD}$ | $Train_{AD}$+ $RT_{B1}$ | $Train_{AD}$+ $RT_{B1}$ |
| Test data | $D_{B2}$ | $D_{B2}$ | $D_{B2}$ |
| Accuracy(%) | 76.9 | 78.1 | 80.3 |

Note: user A and user D are known users, user B is a new user

**Table 13. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{AD}$ | $Train_{AD}$+ $RT_{B1}$ | $Train_{AD}$+ $RT_{B1}$ |
| Test data | $Test_{AD}$ | $Test_{AD}$ | $Test_{AD}$ |
| Accuracy(%) | 99 | 98.7 | 98.7 |

Note: user A and user D are known users, user B is a new user

**Table 14. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BD}$ | $Train_{BD}$+ $RT_{A1}$ | $Train_{BD}$+ $RT_{A1}$ |
| Test data | $D_{A2}$ | $D_{A2}$ | $D_{A2}$ |
| Accuracy(%) | 84.8 | 86.3 | 86.1 |

Note: user B and user D are known users, user A is a new user

**Table 15. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BD}$ | $Train_{BD}$+ $RT_{A1}$ | $Train_{BD}$+ $RT_{A1}$ |
| Test data | $Test_{BD}$ | $Test_{BD}$ | $Test_{BD}$ |
| Accuracy(%) | 97.9 | 97.9 | 97.8 |

Note: user B and user D are known users, user A is a new user

**Table 16. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Train data | $Train_{BC}$ | $Train_{BC}+RT_{D1}$ | $Train_{BC}+ RT_{D1}$ |
| Test data | $D_{D2}$ | $D_{D2}$ | $D_{D2}$ |
| Accuracy(%) | 89.7 | 90.6 | 92 |

Note: user B and user C are known users, user D is a new user

**Table 17. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BC}$ | $Train_{BC}+ RT_{D1}$ | $Train_{BC}+ RT_{D1}$ |
| Test data | $Test_{BC}$ | $Test_{BC}$ | $Test_{BC}$ |
| Accuracy(%) | 98.1 | 97.9 | 97.9 |

Note: user B and user C are known users, user D is a new user

**Table 18. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BD}$ | $Train_{BD}+RT_{C1}$ | $Train_{BD}+ RT_{C1}$ |
| Test data | $D_{C2}$ | $D_{C2}$ | $D_{C2}$ |
| Accuracy(%) | 88.4 | 90.4 | 91.5 |

Note: user B and user D are known users, user C is a new user

**Table 19. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known Users**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{BD}$ | $Train_{BD}+ RT_{C1}$ | $Train_{BD}+ RT_{C1}$ |
| Test data | $Test_{BD}$ | $Test_{BD}$ | $Test_{BD}$ |
| Accuracy(%) | 98.2 | 97.6 | 97.9 |

Note: user B and user D are known users, user C is a new user

**Table 20. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on New User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{CD}$ | $Train_{CD}+RT_{B1}$ | $Train_{CD}+ RT_{B1}$ |
| Test data | $D_{B2}$ | $D_{B2}$ | $D_{B2}$ |
| Accuracy(%) | 81.6 | 83.5 | 84.2 |

Note: user C and user D are known users, user B is a new user

**Table 21. The Recognition Performance Comparison of Deng's Approach and Proposed Approach on Known User**

| | original ELM | Deng's approach | proposed approach |
|---|---|---|---|
| Train data | $Train_{CD}$ | $Train_{CD}+ RT_{B1}$ | $Train_{CD}+ RT_{B1}$ |
| Test data | $Test_{CD}$ | $Test_{CD}$ | $Test_{CD}$ |
| Accuracy(%) | 98.4 | 97.9 | 98.2 |

Note: user C and user D are known users, user B is a new user

## 5. Conclusion and Future Work

In this paper, we proposed an approach USP-ELM to improve the recognition performance. We combined Uncertainty Sampling strategies and posterior probability

 

ELM to solve the problems of common misclassification and how to effectively apply the model trained by known users to new users. We transferred the actual outputs of the ELM model to posterior probabilities, and used Least Confidence and Margin sampling strategy to adapt the ELM model in order to improve the performance of the recognition system. Experiment results show the efficiency of the proposed approach.

The confidence level thresholds in the proposed approach were manually set, how to set an appropriate threshold was very time-consuming, and it depended on the experience of different researchers. In the future, we will research how to set the threshold automatically, and how to combine temporal model such as Hidden Markov Model with posterior probability ELM to correct the misclassification in a sequential stream sensor data.

## 6. Acknowledgments

This work is supported by National Project (No. 61332017 and 2015BAK04B05), National Key Research and Development Program (No.2016YFB1001404), and National Natural Science Foundation of China (Key Project No.61432004), thanks for financial aid for the research.

## References

[1] E. Kim, S. Helal and D. Cook, "Human activity recognition and pattern discovery", IEEE Pervasive Computing, vol. 9, no. 1, **(2010)**, pp. 48-53.
[2] Y. Chen, Z. Zhao, S. Wang and Z. Chen, "Extreme learning machine-based device displacement free activity recognition model", Soft Computing, vol. 16, no. 9, **(2012)**, pp. 1617-1625.
[3] W. Y. Deng, Q. H. Zheng and Z. M. Wang, "Cross-person activity recognition using reduced kernel extreme learning machine", Neural Networks, vol. 53, no. 5, **(2014)**, pp. 1-7.
[4] J. L. R. Ortiz, "Smartphone-based human activity recognition", Springer, **(2015)**.
[5] J. L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra and D. Anguita, "Transition-aware human activity recognition using smartphones", Neurocomputing, vol. 171, no. 1, **(2016)**, pp. 754-767.
[6] J. L. Reyes-Ortiz, L. Oneto, A. Ghio, A. Samá, D. Anguita and X. Parra, "Human activity recognition on smartphones with awareness of basic activities and postural transitions", International Conference on Artificial Neural Networks, **(2014)** September.
[7] Ö. Yürür, C. H. Liu, Z. Sheng, V. C. Leung, W. Moreno and K. K. Leung, "Context-Awareness for Mobile Sensing: A Survey and Future Directions", IEEE Communications Surveys & Tutorials, vol. 18, no. 1, **(2016)**, pp. 68-93.
[8] B. Settles, "Active learning literature survey", University of Wisconsin, Madison, vol. 39, no. 2, **(2010)**, pp. 127-131.
[9] W. Xiao and Y. Lu, "Daily human physical activity recognition based on kernel discriminant analysis and extreme learning machine", Mathematical Problems in Engineering, vol. 2015, **(2015)**, pp. 1-8.
[10] W. He, Y. Guo, C. Gao and X. Li, "Recognition of human activities with wearable sensors", EURASIP Journal on Advances in Signal Processing, vol. 1, **(2012)**, pp. 1-13.
[11] Y. Zheng, "Human activity recognition based on the hierarchical feature selection and classification framework", Journal of Electrical and Computer Engineering, vol. 2015, no. 34. **(2015)**, pp. 1-9.
[12] H. Yu, C. Sun, W. Yang, X. Yang and X. Zuo, "AL-ELM: One uncertainty-based active learning algorithm using extreme learning machine", Neurocomputing, vol. 166, **(2015)**, pp. 140-150.
[13] J. Liu, H. Yu, W. Yang and C. Sun, "Combining Active Learning and Semi-Supervised Learning Based on Extreme Learning Machine for Multi-class Image Classification", International Conference on Intelligent Science and Big Data Engineering, **(2015)**, pp. 163-175.
[14] G. B. Huang, Q. Y. Zhu and C. K. Siew, "Extreme learning machine: theory and applications", Neurocomputing, vol. 70, no. 1, **(2006)**, pp. 489-501.
[15] G. B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme learning machine for regression and multiclass classification", Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 42, no. 2, **(2012)**, pp. 513-529.
[16] Z. Yong and J. E. Meng, "Sequential active learning using meta-cognitive extreme learning machine", Neurocomputing, vol. 173, no. 3, **(2016)**, pp. 835-844.
[17] L. Hu, Y. Chen, S. Wang and Z. Chen, "b-COELM: A fast, lightweight and accurate activity recognition model for mini-wearable devices", Pervasive and Mobile Computing, vol. 15, **(2014)**, pp. 200-214.
[18] J. Yang, H. Yu, X. Yang and X. Zuo, "Imbalanced Extreme Learning Machine Based on Probability Density Estimation", International Workshop on Multi-disciplinary Trends in Artificial Intelligence, **(2015)**, pp. 160-167.

[19] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley and B. W. Suter, "The multilayer perceptron as an approximation to a Bayes optimal discriminant function", Neural Networks, IEEE Transactions on, vol. 1, no. 4, **(1990)**, pp. 296-298.
[20] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods", Advances in large margin classifiers, vol. 10, no. 3, **(1999)**, pp. 61-74.
[21] "Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set", [Online]. Available:http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions

## Authors

**Xiaopeng Niu,** he is a Ph.D. student in School of Computer and Communication Engineering, University of Science and Technology Beijing. His research interests include Pervasive Computing, Internet of Things, machine learning, Human-Computer Interaction, Artificial Intelligence.

**Zhiliang Wang,** he is a professor in School of Computer and Communication Engineering, University of Science and Technology Beijing. His research interests include Affective Computing, Internet of Things, and Human-Computer Interaction.

**Wendong Xiao,** he is a professor in School of Automation and Electrical Engineering, University of Science and Technology Beijing. His research interests include Wireless Sensor Network, Internet of Things, Wearable Computing, and Wireless Location.

**Zhigeng Pan,** he is a professor in Digital Media and HCI Center, Hangzhou Normal University. His research interests include Virtual Reality, Human-Computer Interaction, and Digital Media.