A correlation-based binary particle swarm optimization method for feature selection in human activity recognition

Huaijun Wang[1,2], Ruomeng Ke[1], Junhuai Li[1,2], Yang An[1], Kan Wang[1,2] and Lei Yu[1,2]

Abstract

Effective feature selection determines the efficiency and accuracy of a learning process, which is essential in human activity recognition. In existing works, for simplification purposes, feature selection algorithms are mostly based on the assumption of feature independence. However, in some scenarios, the optimization method based on this independence hypothesis results in poor recognition performance. This article proposes a correlation-based binary particle swarm optimization method for feature selection in human activity recognition. In the proposed algorithm, the particle swarm optimization algorithm is no longer used as a black box. Meanwhile, correlation coefficients among the features are added to binary particle swarm optimization as a feature correlation factor to determine the position of particles, so that the feature with more information is more likely to be selected. The *k*-nearest neighbor classifier is then used as the fitness function in the particle swarm optimization to evaluate the performance of the feature subset, that is, feature combination with the highest *k*-nearest neighbor classifier recognition rate would be picked as the eigenvector. Experimental results show that the proposed method can work well with six classifiers, namely, J48, random forest, *k*-nearest neighbor, multilayer perceptron, naïve Bayesian, and support vector machine, and the new algorithm can improve the classification accuracy in the OPPORTUNITY Activity Recognition dataset.

Keywords

Activity recognition, sensor, feature selection, binary particle swarm optimization, feature correlation

Introduction

With the rapid development of ultra-low-power sensor technology, sensor-based applications could facilitate our daily life in various aspects.[1] Compared with vision-based activity recognition, the sensor-based one could better embody the essence of activities. Moreover, thanks to sensors' small volume, high sensitivity, and simple apparatus characteristics, sensor-based activity recognition can preserve users' privacy and acquire data unobtrusively, making it widely applicable in healthcare and elderly care. Therefore, research on sensor-based activity

[1]School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, China
[2]Shaanxi Key Laboratory for Network Computing and Security Technology, Xi'an University of Technology, Xi'an, China

Corresponding author:
Junhuai Li, School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048, China.
Email: lijunhuai@xaut.edu.cn

recognition has been attracting growing attention in a number of disciplines and application domains.[2–4]

Although some works have been done on sensor-based activity in recent years, there are significant problems yet to be solved.[5] In particular, when attempting to recognize human activity with better accuracy, the model could get more complicated as more features are extracted, resulting in curse of dimensionality. By eliminating the irrelevant or redundant ones to reduce the feature number, running time of feature selection could be significantly reduced, while model precision can be improved.[6] Feature selection refers to the selection of a subset with less redundant features and better classification results, that is, to construct a new feature vector with fewer features for activity characterization. As such, the high-dimensional feature turns into a low-dimensional one, enabling efficient reduction of computational complexity.

Common activity features (include time, frequency, and time–frequency domains) could only reflect a behavior from one aspect. Simply combing these features cannot improve the recognition accuracy and would increase the redundancy and weaken the feature expression instead. Feature selection not only enhances the expressive power of the feature category, but also greatly reduces the feature dimension, so as to improve the timeliness of the activity recognition.

To simplify the problem and reduce the runtime, most existing works on feature selection assume that each feature is independent of each other. However, in some scenarios, the optimization method based on the independence hypothesis degrades the recognition performance. For example, a feature may be irrelevant to the target when considered alone, but it may be very relevant to the target if it is considered together with the other features. Different feature combinations have different contributions to the classification results. However, due to correlation among features, the optimal feature subset is typically not a simple combination of multiple individual ones. Therefore, it is necessary to felicitously analyze the correlation between features. In this article, we take the correlation among features into consideration and propose a correlation-based binary particle swarm optimization (BPSO) method for feature selection in human activity recognition.

First, rather than using particle swarm optimization (PSO) as a black box, we attach the correlation among features to the BPSO and deem it as a feature correlation factor to determine the position of particle, so that the feature with a larger amount of data information is more likely to be selected, thereby enhancing the performance of feature selection.

Second, the *k*-nearest neighbor (KNN) classifier is used as the fitness function in the PSO to evaluate the performance of the feature subset, that is, the feature combination with the highest KNN classifier recognition rate would be picked as the eigenvector. This

approach not only reduces the classifier computational complexity, but also improves the timeliness and accuracy of activity recognition.

Finally, extensive experiments are conducted to evaluate the proposed approach. The results show that the proposed feature selection algorithm outperforms the traditional ones and could improve the recognition rate considerably.

Related works

Feature selection is not only the key problem in pattern recognition, but also one of the important issues in sensor-based activity recognition field. Appropriate feature set has important influence on the accuracy of recognition. In general, an activity could be described with better specificity if more categories of features are extracted. However, with the increase of new features, eigenvector dimension extracted by sample data is growing, leading to a rapid increase in time consumption and computational complexity. Therefore, it is necessary to select a feature set that contains the least number of feature dimensions and contributes to higher recognition accuracy. Feature selection can eliminate irrelevant or redundant features, thus improving the accuracy and reducing the computational complexity of the model.

According to whether the feature subset evaluation criteria are related to the classification algorithm, feature selection can be divided into three categories: filter based, wrapper based, and embedded approach based.[7–9] The filter-based methods separate the feature selection process from the classification verification process, so that the feature selection process is independent of the classifier. In addition, the evaluation criteria do not depend on the specific classifier, but rely on the attribute information of the data feature itself only. Therefore, the methods using this approach are typically fast but they need a threshold as the stopping criterion for feature selection. Several filter-based methods have been proposed in the literature including information gain,[10] gain ratio,[11] term variance,[12] Gini index,[13] Laplacian score,[14] Fisher score,[15] minimal-redundancy-maximal-relevance,[16] random subspace method,[17] relevance-redundancy feature selection,[18] unsupervised feature selection method based on ant colony optimization (UFSACO),[19] relevance–redundancy feature selection based on ant colony optimization (RRFSACO),[20] graph clustering with node centrality for feature selection method (GCNC),[21] and graph clustering based ant colony optimization feature selection method (GCACO).[22]

Wrapper-based methods combine feature selection with the design of the classifier and evaluate the feature subsets on the basis of the accuracy of classification. Examples of wrapper-based methods include sequential backward selection, sequential forward selection,[23] ant

colony optimization,[24] PSO,[25] genetic algorithm (GA),[26,27] random mutation hill-climbing,[28,29] simulated annealing,[30] and artificial bee colony.[31,32] Although these approaches could generally achieve a better result compared to the filter-based approach, they require more computational resources in order to select the best feature subset.

The combination of filter- and wrapper-based approaches produces the embedded approach–based method. Embedded approach–based methods seek to subsume feature selection as part of the model building process and are associated with a specific learning model. First, the filter-based approach is used to eliminate the irrelevant features and noise features, so as to reduce the number of features to be selected. On the basis of that, the wrapper-based approach is applied to optimize the feature set. The embedded approach–based method integrates the high efficiency of the filter model and the high accuracy of the wrapper model, therefore achieving a better feature subset.[33–35]

Considering feature selection into the human activity recognition chain can improve performance than the one achieved when the entire feature set is used.[36] S Chernbumroong et al.[37] proposed a feature selection for multi-sensor activity recognition based on maximum relevancy maximum complementary. This method is different from the traditional ones using correlation and redundancy as the selection criterion and takes into account the complementarity among features, thereby improving the recognition rate of multi-sensor activity. H Fang et al.[38] integrated the back-propagation neural network algorithm with the principle of distance maximization between classes for feature selection. N Oukrich et al.[39] used multilayer perception and back-propagation algorithm to train the neural network, and selected the features in line with the minimum redundancy maximum correlation criterion to recognize activity behavior. Moreover, J Suto et al.[40] compared wrapper feature selection with the filter-based method, and their experimental results showed that the former one outperformed the latter one in terms of activity recognition. MT Uddin and MA Uddiny[41] proposed a guided random forest-based feature selection algorithm and leveraged the activity dataset to train the random forest to distinguish the importance between different features. Furthermore, multi-attribute fusion was used in Wang and Huo[42] to jointly incorporate the mutual information, interclass difference information, intraclass fluctuation information, and computational complexity. In order to select important discriminating features to recognize the human activities, the features were selected based on spatiotemporal orientation energy and template matching, and the relevant features were identified by gradient boosting and random forest.[43] S González et al.[44] first analyzed the information correlation coefficient and then adopted the wrapper method for feature selection. UM Nunes et al.[45] presented a novel framework using max-min features and key poses with differential evolution random forest classifier, which had no thresholds to tune. In addition, X Xian et al.[46] proposed a modified method with linear discriminant analysis (LDA) based on GA, and their testing results showed that the modified method can effectively raise the recognition accuracy.

In existing works on activity recognition, feature selection algorithms are mostly based on the assumption of feature independence. The independence hypothesis simplifies the model to a certain extent and reduces the selection time, but results in poor recognition results in some scenarios due to interdependence among features. For instance, if a feature is considered separately, it may not be related to the target object, but it may be very relevant to the target if it is considered together with other features.

Correlation-based BPSO method for feature selection

BPSO

PSO is a meta-heuristic search algorithm proposed by Drs Eberhart and Kennedy in 1995. The algorithm originated from the behavior of bird predation, which is simple and easy to implement, and is the most widely used swarm intelligence optimization algorithm.[47,48] The basic idea behind PSO is to find the optimal solution through mutual cooperation and information sharing among individuals in a swarm. Inspired by the regularity of the birds' cluster activities, it is actually a simplified model on the basis of swarm intelligence. When observing the activity of animal groups, the PSO changes the motion direction of the whole group from disorder to order in the feasible solution space, according to information sharing between individuals. For illustration, consider a group of birds (particle swarms) randomly searching for food in a region (a solution of the problem), with only one objective (optimal solution). All birds do not know the explicit position of the food at first but are aware of the distance between themselves (the current optimal solution of the individual) and food (the fitness of the solution). So the most intuitive way to find the target food is to search over the surrounding area of the nearest bird (the optimal solution of the group). In particular, each bird in the flock, based on the optimal position of the current flock and the best position it has flown over, constantly adjusts its current position and velocity to search for food.

More specially, Kennedy and Eberhart[49] proposed BPSO to solve the discrete optimization problem in 1997. For BPSO, the velocity vector of a particle is no longer the change rate of the particle position, but the

probability of change for that particular particle position. Depending on the velocity, the value of the particle is set to be 1 or 0. The solution processes in the *m*-dimensional space of the BPSO are as follows:

1. Randomly generate *n* *m*-dimensional particles, with the position $X_i = (x_{i1}, x_{i2}, \ldots, x_{im})^T$, and the velocity $V_i = (v_{i1}, v_{i2}, \ldots, v_{im})^T (i = 1, 2, \ldots, n)$.
2. Use the fitness function $F(x)$ to calculate the fitness for each particle, and find the optimal solution g_{best} in the group and individual optimal solution p_{best} for each individual.
3. Determine whether the convergence conditions are met: if met, go to step (5); otherwise, go to the next step.
4. Each particle tracks p_{best} and g_{best} to update its velocity and position by equations (1) and (2)

$$v_{im}^{k+1} = w v_{im}^k + c_1 r_1^k (p_{Best\_im}^k - x_{im}^k) + c_2 r_2^k (g_{Best\_m}^k - x_{im}^k) \tag{1}$$

$$\begin{cases} x_{im}^{k+1} = 1, & r_{im}^{k+1} < sigmoid(v_{im}^{k+1}) \\ x_{im}^{k+1} = 0, & r_{im}^{k+1} \geq sigmoid(v_{im}^{k+1}) \end{cases},$$
$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

where $k \in n$ is the iteration index, v_{im} is the velocity of the *i*th particle, and x_{im} is the position of the *i*th particle at the *d*th dimension; *w* denotes the inertia weight, c_1 and c_2 are the acceleration constants, r_1, r_2, and r_{im} are the random numbers generated in the interval $[0, 1]$, p_{Best_im} is the optimum position of the *i*th particle in the *m*th dimension, and g_{Best_m} is the best position for the entire particle swarm in the *m*th dimension.

5. After the update is completed, go to step (2).
6. End.

Feature selection algorithm

PSO algorithm is easy to understand and simple to implement, and can solve many optimization problems. Therefore, studies on PSO for feature selection have concerned experts and scholars from both academia and industry. Although PSO has good optimization performance, it still has a lot of room for improvement when it comes to specific optimization problems. Most existing feature selection algorithms based on PSO ignore the particularity of feature selection and only take advantage of the search ability of particle swarm. However, if the differences between fitness functions are neglected, the existing feature selection by PSO is almost identical to other optimization problems.

The traditional feature selection typically assumes that the feature is independent of each other. However, there exists the correlation between features, and the optimal subset is typically not a simple combination of multiple individual optimal features, thus necessitating the analysis of the correlation between features. Moreover, the studies on feature selection mostly ignore the particularity of feature selection, but simply use the search ability of the PSO algorithm. At this time, PSO is only deemed as a black box.[50] Based on the above considerations, this article focuses on the combination of feature optimization and PSO, and proposes a feature correlation–based method to improve the search efficiency of PSO. This method increases the probability of feature with more information being selected and effectively selects features with larger discriminative power between classes, so as to reduce high computational cost and improve poorer performance of classification algorithms caused by redundant features. It also improves the timeliness and accuracy of activity recognition. The location update strategy adopted by traditional BPSO algorithm lacks the consideration of correlation between features, which makes it unsuitable for feature optimization problems. In this article, a new location update strategy for feature selection is designed to improve the optimization ability of BPSO. Intuitively, in each component of the position vector, the new position update strategy takes overall consideration of the particle's current velocity, current position, the best position experienced by each individual, the optimal location of the population, and the relevance between features to be selected. As such, the feature with a larger amount of data would be selected with higher probability, resulting in a considerable improvement in optimization performance. Next, we will elaborate our proposed feature selection method–based BPSO with consideration of feature correlation in human activity recognition.

Given a *p*-dimensional feature set $F = \{f_1, f_2, \ldots, f_p\}$, the correlation coefficient between features in *F* is calculated first in the initializing stage as follows

$$r_{ij} = \frac{\sum_{k=1}^{n} f_i(k) f_j(k)}{\left[\sum_{k=1}^{n} f_i^2(k) \sum_{k=1}^{n} f_j^2(k)\right]^{\frac{1}{2}}} \tag{3}$$

From the Cauchy–Schwartz inequality, it follows that the absolute value of the correlation coefficient is not greater than 1. A larger r_{ij} indicates a larger similarity between f_i and f_j, that is, a greater likelihood of belonging to the redundant features. From equation (3), we could acquire the similarity matrix between features in *F* as follows

Pseudo-code of correlation-based BPSO method for feature selection
1	**Input** F: original feature set
2	N: size of population
3	D: dimension of feature
4	**Output** S: optimal feature subset
5	**Initialize** each particle in the population
6	**Calculate** the matrix R of correlation coefficients between features in F
7	**Calculate** the contribution of each feature in F by R
8	**while** The termination condition of the iteration is not satisfied **do**
9	**for** i=1 to N **do**
10	**Calculate** the fitness value of the particle using KNN classifier
11	**Update** the historical best position of the particle
12	**end for**
13	**Update** the optimal position of the population
14	**for** i=1 to N **do**
15	**for** i=1 to D **do**
16	**Update** the velocity of the particle
17	**Update** the position of the particles combining the w value of each feature
18	**end for**
19	**end for**
20	**end while**
21	**Output** the optimal position　（optimal feature subset）

Figure 1. Pseudocode of the proposed feature selection algorithm.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (4)$$

where $r_{ii} = 1$, $i \in [1,p]$, holds since f_i is positively related to itself. For two known sample sets, the similarity measure between the feature f_i and the other features is as follows

$$q_i = \frac{\sum_{j=1}^{n} |r_{ij}|}{p} \quad (5)$$

which could be utilized to obtain the similarity measure of all the features

$$q = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} \quad (6)$$

Considering the similarity between features (i.e. redundancy), q_i is utilized to evaluate the contributions of distinguishability. The larger the q_i, the greater the redundancy as well as the smaller the weight, and vice versa. Thus, we have the following equation

$$w_i = 1 - q_i \quad (7)$$

Typically, w_i is normalized such that $\sum_{i=1}^{n} w_i = 1$ holds. As mentioned above, we could calculate the weight of each feature and further take it as one part of particle position determination. As such, the feature with a larger amount of data would be selected more likely and equation (2) can be reformulated as equation (8), that is

$$\begin{cases} x_{im}^{k+1} = 1, & r_{im}^{k+1} + w_{iavg} < sigmoid(v_{im}^{k+1}) + w_{im} \\ x_{im}^{k+1} = 0, & r_{im}^{k+1} + w_{iavg} \geq sigmoid(v_{im}^{k+1}) + w_{im} \end{cases},$$
$$sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (8)$$

Table 1. S-shaped and V-shaped families of transfer functions.

S-shaped family		V-shaped family	
Name	Transfer function	Name	Transfer function
S1	$T(x) = \frac{1}{1 + e^{-2x}}$	V1	$T(X) = \left\| erf\left(\frac{\sqrt{\pi}}{2}x\right) \right\| = \left\| \frac{\sqrt{2}}{\pi} \int_{0}^{(\sqrt{\pi}/2)x} e^{-t^2} dt \right\|$
S2	$T(x) = \frac{1}{1 + e^{-x}}$	V2	$T(x) = \|\tanh(x)\|$
S3	$T(x) = \frac{1}{1 + e^{(-x/2)}}$	V3	$T(x) = \left\| \frac{(x)}{\sqrt{1 + x^2}} \right\|$
S4	$T(x) = \frac{1}{1 + e^{(-x/3)}}$	V4	$T(x) = \left\| \frac{2}{\pi}\arctan\left(\frac{\pi}{2}x\right) \right\|$

Table 2. The number of activity samples.

Activity	Tag	Size
Standing	1	155,765
Walking	2	80,474
Sitting	4	48,869
Lying	5	5572
Total		290,680

where w_{iavg} represents the averaged contribution over all features in the ith particle and w_{im} denotes the contribution of the mth dimension in the ith particle.

The traditional PSO mainly utilizes real-number encoding to solve the continuous optimization problem; nevertheless, most selection problems are only with discrete data, thus constraining the PSO to some extent. In view of the particularity of feature selection, binary encoding is adopted, which encodes each dimension feature by binary numbers.

Meanwhile, to enable a higher recognition rate for the selected feature subset, we utilize wrapper feature selection. In other words, the fitness function of the PSO is a certain classification algorithm. KNN classification algorithm is simple and easy to implement, and supports incremental learning. In this study, the KNN classifier is chosen as the fitness function in the PSO to evaluate the performance of a feature subset. Then the feature set with the highest recognition rate of KNN classifier is selected as the final feature vector.

The pseudocode is shown in Figure 1, which starts with the initialization of each particle in the population. Then the correlation coefficients between features and the contribution of each feature in all feature sets are calculated to obtain the optimal initial value of the individual and the population, as defined by lines 6 and 7. The search process will stop and output the optimal position of the population, if the process comes to a stopping condition; otherwise, it will execute circularly. Two stopping conditions are defined—either when the error reaches the preset requirement or when the number of iterations exceeds the maximum allowable

number of iterations. Lines 8–20 define the detailed cyclic process to get the optimal feature subset, which contains updating the particles' velocity and position, calculating the fitness value of the particle using KNN classifier, and updating the historical best position of the particle and the optimal position of the population. The optimal feature subset will be selected after finishing this cyclic process.

PSO is an optimization method and has good local search ability. Therefore, it is a hot research topic to use PSO to solve the problem of feature selection. There are a number of publications that have reviewed the work done so far in this area; for example, the BPSO is used to select effective text features in text data and extract effective physiological signal to classify the emotional state, and in addition it is also widely used in face recognition. So the proposed approach can be easily extended to different problems of feature selection. The main part of this algorithm is a transfer function which is for mapping a continuous search space to a discrete search space. In the study of Mirjalili and Lewis,[51] the sigmoid function in equation (8) has eight alternatives, including four S-shaped functions and four V-shaped functions, as shown in Table 1. Therefore, in this experiment, the eight transformation functions are used for feature optimization, and the results are compared.

Experimental results and discussion

Here, we show the effectiveness of the proposed method with the experimental results, where the OPPORTUNITY Activity Recognition dataset[52,53] is employed. In particular, the acceleration sensor data of positions 1, 2, and 3 in the dataset are utilized to analyze four basic activities (i.e. standing, walking, sitting, and lying), as illustrated in Figure 2. The total number of raw data is 358,987, while that of the preprocessed data turns out to be 290,680, with a detailed volume list (as shown in Table 2) for each specific activity.

In the experiment, the BPSO iterates 500 times. The number of initial particles is 30 and the particle
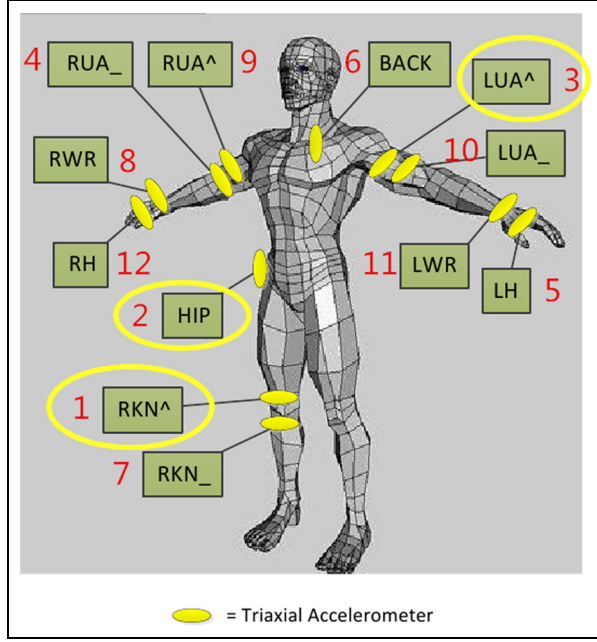
Figure 2. Sensor deployment of the OPPORTUNITY dataset.

dimension is identical to the feature dimension. Majority of the approaches used window size in the range of 0.1–10 s, with an overlapping either absent or limited to 50% between consecutive windows.[54,55] Banos et al.[56] proved that short windows normally resulted in better recognition performance, and the interval of 1–2 s would exhibit the best trade-off between recognition speed and accuracy from a global perspective. Therefore, the window size is set to be 0.5 (32 pieces of data in a window), 1 (64 pieces of data in a window), and 2 s (128 pieces of data in a window) for data partitioning and feature extraction, respectively. Meanwhile, a total of 11 features, that is, mean, variance, root mean square, mean absolute deviation, range, covariance, quartile deviation, coefficient of correlation, kurtosis, skewness, and energy, are utilized for selection, as shown in Table 3.

Taking into account two factors, that is, sensors in different positions and different axes in the same sensor, the selection experiment can be categorized into two schemes: (1) considering only the influence of sensors' positions on feature selection results, excluding the possible effect of different sensor axes, with a 33-dimensional feature set, and (2) involving only the effect of each feature in the same sensor on feature selection, without considering the influence of different axes of sensor and the different axes of the same sensor. Also, feature selection for sensor data is with three positions, that is, RKN, HIP, and LUA, respectively, and the feature set for each position is 10-dimensional (correlation coefficient is not taken into account). For each scheme, the experiment is further divided into two steps. First of all, pairwise combinations of all four behaviors form six types of activity groups, that is, sitting–lying, walking–lying, standing–walking, standing–lying, walking–sitting, and standing–sitting. By feature selection experiments, we can select the feature subsets which can be utilized to effectively distinguish among these activity groups and then analyze the optimal distinguishing feature sets of pairwise behaviors in each group (the experiment shows that the size of the window is 32 s and the feature dimension is 33, and the size of the other windows and the feature dimensions are similar to those of the group). Then, the four behaviors are assembled into an activity group for feature selection, and features that could effectively distinguish among all four activities are selected.

The 33 types of features are listed in Table 4. For the case with a window size of 0.5 s, the number of feature samples is set to be 18,162 and scheme 1 is conducted. The experimental results of four activities in pairwise combinations are shown in Tables 5–10 and Figure 3, and those of the four behaviors as a behavior group are presented in Table 11 and Figure 4.

From the experimental results of the pairwise combinations (as shown in Table 11), it follows that for all six combinations of activities the feature selection proposed in this article can identically select the feature combination with lower dimension and higher recognition rate. Among all six classification algorithms, that is, J48, random forest (RF), KNN, multilayer perceptron (MLP), naïve Bayesian (NB), and support vector machine (SVM), the SVM and NB classifiers combined with our proposed method demonstrate the best recognition effect. It could not only greatly reduce the feature dimension, but also improve the recognition rate. Meanwhile, J48 and RF are suboptimal, while KNN and MLP recognition produce the worst performance. In terms of the overall recognition effect, the recognition rates of the RF and MLP classifiers are the highest, which are 99.74% and 99.83%, respectively. In addition, as compared to J48 and KNN, NB and SVM show lower recognition rates. It should be noted that the sensor data of RKN an LUA contribute significantly to distinguishing between sitting–lying, walking–sitting, and standing–sitting. For distinguishing between walking and lying, the contributions from the sensor data of HIP and LUA are greater. Moreover, the sensor data of RKN, HIP, and LUA are efficient for distinguishing between standing and walking. For the distinction between standing and lying, RKN and HIP contribute more.

It can be seen from Table 11 that the recognition rate of walking–lying is the highest among all six action combinations, which reaches 99.65% when using J48

Table 3. Features used in experiment.

Feature name	Notation	Equation
Mean	M	$M = \frac{1}{N}\sum_{i=1}^{N} x_i$
Variance	V	$V = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$
Root mean square	RMS	$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Mean absolute deviation	MAD	$MAD = \frac{1}{N}\sum_{i=1}^{N}|x_i - \bar{x}|$
Range	R	$R = \max - \min$
Covariance	COV	$COV(X,Y) = E[(X - E[X])(Y - E[Y])]$
Quartile deviation	QD	$Q_1 = \frac{(N+1)}{4},\ Q_2 = \frac{3(N+1)}{4}$
Coefficient of correlation	CORR	$CORR(x,y) = \dfrac{\sum_{i=1}^{N}[(x_i-\bar{x})(y_i-\bar{y})]}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2 \times \sum_{i=1}^{N}(y_i-\bar{y})^2}}$
Kurtosis	KS	$KS = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^4$
Skewness	SS	$SS = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma}\right)^3$
Energy	E	$E = \frac{1}{N}\sum_{i=1}^{N}|F(x_i)|^2$

Table 4. Feature types.

Number	Name	Number	Name
1	RKN mean	18	LUA range
2	HIP mean	19	RKN kurtosis
3	LUA mean	20	HIP kurtosis
4	RKN variance	21	LUA kurtosis
5	HIP variance	22	RKN skewness
6	LUA variance	23	HIP skewness
7	RKN mean absolute deviation	24	LUA skewness
8	HIP mean absolute deviation	25	RKN covariance
9	LUA mean absolute deviation	26	HIP covariance
10	RKN root mean square	27	LUA covariance
11	HIP root mean square	28	RKN–HIP coefficient of correlation
12	LUA root mean square	29	HIP–LUA coefficient of correlation
13	RKN quartile deviation	30	LUA–RKN coefficient of correlation
14	HIP quartile deviation	31	RKN energy
15	LUA quartile deviation	32	HIP energy
16	RKN range	33	LUA energy
17	HIP range		

and MLP classifiers. It should be noted that the distinction degree of lying is higher than those of the other three actions. Walking–sitting and standing–sitting have medium recognition rates, which shows that the sitting action is with moderate distinction. The activity with the lowest recognition rate is standing–walking, while the highest recognition rate of the six classifiers reaches 88.39%, indicating that the distinction between standing and walking is relatively low. It could be concluded that J48, RF, and MLP have the best recognition effect, while NB and SVM are of inferior effect and KNN is of medium effect.

The window size is 0.5 s and the number of feature samples is 18,162. The results of experiment carried out according to scheme 1 are shown in Table 12 and Figure 4, while Table 13 shows the results of scheme 2.

Table 5. Recognition rate and selected feature subset in the case of sitting–lying.

Transfer function		Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	18	{2,3,6,8,9,10,12,14,19,21,22,23,24,26,27,28,29,30}	99.29	99.41	97.26	99.47	95.76	89.82
	S2	17	{2,3,8,10,11,12,13,15,19,20,21,22,23,24,28,29,30}	99.29	99.53	98.18	99.26	98.21	96.65
	S3	16	{1,4,6,8,9,13,14,16,21,23,25,27,28,29,31,33}	99.00	99.26	98.82	99.44	98.53	89.82
	S4	18	{1,2,6,7,8,11,13,19,20,21,22,23,24,26,27,28,29,30}	99.15	99.32	98.59	99.24	98.47	89.82
V-shaped	V1	9	{1,3,12,13,15,20,21,22,30}	99.15	99.38	98.88	99.41	99.41	96.82
	V2	9	{1,9,10,12,13,20,24,28,30}	99.21	99.18	98.71	99.18	98.88	97.06
	V3	11	{1,3,9,12,15,19,20,21,23,24,30}	99.21	99.18	98.00	99.09	98.12	97.18
	V4	11	{1,7,9,12,13,15,19,20,21,24,30}	98.94	99.26	98.62	99.44	98.18	96.85
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20, 21,22,23,24,25,26,27,28,29,30,31,32,33}	99.09	99.38	99.18	99.38	98.62	89.92
Optimal subset	V1	9	{1,3,12,13,15,20,21,22,30}	99.15	99.38	98.88	99.41	98.41	96.82

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 6. Recognition rate and selected feature subset in the case of walking–lying.

Transfer function		Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	22	{1,3,5,6,7,10,11,12,15,16,17,19,20,22,24,25, 26,27,28,29,31,32}	99.57	99.61	99.31	99.72	97.95	93.56
	S2	13	{1,2,11,14,15,17,19,20,21,22,23,28,29}	99.67	99.70	99.13	99.65	99.29	97.26
	S3	7	{1,8,14,15,21,23,29}	99.65	99.61	99.37	99.65	99.20	98.38
	S4	15	{2,7,8,10,15,17,19,20,21,22,23,24,28,29,30}	99.53	99.74	98.46	99.72	98.23	97.73
V-shaped	V1	14	{6,7,11,12,13,14,21,22,24,26,28,29,31,32}	99.63	99.74	98.46	99.53	96.93	93.56
	V2	16	{1,3,4,5,6,7,9,13,14,18,27,28,29,30,31,32}	99.59	99.70	99.05	99.81	97.15	93.56
	V3	13	{3,7,8,11,12,13,15,16,19,23,24,28,31}	99.53	99.63	98.70	99.83	97.43	94.70
	V4	10	{2,8,9,11,13,16,19,22,30,31}	99.65	99.65	99.00	99.68	98.16	95.59
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19, 20,21,22,23,24,25,26,27,28,29,30,31,32,33}	99.63	99.70	99.09	99.81	97.19	93.56
Optimal subset	S3	7	{1,8,14,15,21,23,29}	99.65	99.61	99.37	99.65	99.20	98.38

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 7. Recognition rate and selected feature subset in the case of standing–walking.

Transfer function		Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	22	{2,3,6,7,8,12,13,15,16,17,20,21,22,23,24, 26,27,28,29,30,31,32}	87.08	88.23	87.24	87.99	84.91	65.94
	S2	18	{3,6,7,9,11,13,15,16,17,19,22,23,24,26,27,28,31,32}	87.26	87.87	86.84	87.91	85.19	65.94
	S3	16	{1,2,6,11,12,13,16,17,19,24,26,27,28,30,31,32}	87.28	87.74	87.01	88.15	85.26	65.94
	S4	15	{6,7,9,11,15,16,17,19,24,26,27,29,30,31,32}	87.22	88.03	86.78	87.86	84.63	65.94
V-shaped	V1	19	{3,6,7,9,10,11,12,16,17,18,20,21,23,26,27,29,30,31,32}	87.47	88.21	87.05	87.88	84.97	65.94
	V2	18	{3,6,8,9,10,12,13,16,17,18,20,21,23,26,27,28,31,32}	87.68	88.39	87.05	88.37	84.90	65.94
	V3	20	{1,6,8,9,10,11,14,15,16,18,20,21,22,24,26,27,29,30,31,32}	87.30	88.04	86.88	87.74	84.75	65.94
	V4	17	{1,2,3,6,11,14,16,17,18,19,20,21,26,27,29,31,32}	87.18	88.27	86.70	88.27	84.47	65.94
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21, 22,23,24,25,26,27,28,29,30,31,32,33}	88.22	88.23	87.54	88.10	84.93	65.94
Optimal subset	V2	18	{3,6,8,9,10,12,13,16,17,18,20,21,23,26,27,28,31,32}	87.68	88.39	87.05	88.37	84.90	65.94

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 8. Recognition rate and selected feature subset in the case of standing–lying.

Transfer function	Dimension		Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	21	{1,2,4,5,8,9,10,12,13,15,18,19,20,21,23,24,25,26,27,29,31}	99.52	99.73	99.24	99.74	94.33	96.57
	S2	21	{1,2,4,5,7,11,12,13,14,19,20,21,22,23,24,25,26,27,28,29,31}	99.49	99.60	98.77	99.58	94.54	96.57
	S3	19	{2,4,5,7,10,11,12,16,18,19,20,21,23,24,25,27,29,30,31}	99.54	99.65	98.51	99.57	94.58	96.57
	S4	21	{1,2,4,5,7,8,9,12,13,14,15,16,19,20,21,22,24,25,28,30,31}	99.52	99.64	98.72	99.61	94.13	96.57
V-shaped	V1	11	{1,8,10,16,20,22,23,25,26,30,31}	99.30	99.38	98.63	99.05	95.53	96.57
	V2	10	{2,7,14,17,19,23,28,29,30,31}	99.57	99.62	98.04	99.46	96.56	97.86
	V3	11	{7,11,13,15,17,19,23,24,28,30,31}	99.52	99.64	97.96	99.48	96.39	97.74
	V4	9	{2,9,11,12,13,14,16,26,31}	99.48	99.66	99.62	99.59	94.17	96.57
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22, 23,24,25,26,27,28,29,30,31,32,33}	99.51	99.70	98.93	99.67	92.40	96.57
Optimal subset	V2	10	{2,7,14,17,19,23,28,29,30,31}	99.57	99.62	98.04	99.46	96.56	97.86

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 9. Recognition rate and selected feature subset in the case of walking–sitting.

Transfer function	Dimension		Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	16	{3,4,11,12,13,16,17,18,19,21,25,28,29,30,31,33}	96.34	96.98	94.73	96.05	90.32	62.23
	S2	20	{2,6,7,8,10,11,12,13,15,16,19,20,21,22,24,25,26,30,31,33}	96.75	97.24	94.98	96.23	90.72	62.23
	S3	16	{2,4,7,8,10,19,20,21,22,23,24,25,27,29,31,33}	96.46	97.09	95.05	96.63	90.71	62.25
	S4	17	{2,3,6,7,8,9,11,12,21,22,23,24,25,26,28,31,33}	96.45	97.28	95.16	96.16	88.85	62.23
V-shaped	V1	19	{3,4,7,8,10,11,12,13,14,15,17,20,23,25,27,28,30,31,33}	96.76	97.07	95.21	96.29	89.73	62.23
	V2	15	{2,4,8,9,10,12,13,16,17,21,22,27,30,31,33}	96.58	97.23	95.71	96.63	90.57	62.23
	V3	18	{1,2,6,8,10,11,12,15,17,18,20,21,23,24,25,26,31,33}	96.62	97.38	96.13	96.66	89.41	62.23
	V4	17	{4,8,9,11,13,15,16,17,18,20,21,25,28,29,30,31,33}	96.75	97.28	94.98	95.89	89.76	62.23
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21, 22,23,24,25,26,27,28,29,30,31,32,33}	96.49	97.20	95.72	96.67	89.99	62.23
Optimal subset	V2	15	{2,4,8,9,10,12,13,16,17,21,22,27,30,31,33}	96.58	97.23	95.71	96.63	90.57	62.23

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 10. Recognition rate and selected feature subset in the case of standing–sitting.

Transfer function	Dimension		Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	17	{1,2,4,10,12,13,20,21,22,23,24,25,26, 29,30,31,33}	90.65	92.59	79.18	90.75	52.03	76.13
	S2	17	{2,3,4,12,13,19,20,21,23,24,25,26,28,29, 30,31,33}	90.61	92.66	73.49	87.71	51.26	76.13
	S3	17	{3,4,10,11,12,19,21,22,23,24,25,26,28, 29,30,31,33}	90.55	92.48	73.90	90.22	52.42	76.13
	S4	16	{3,4,11,12,19,21,22,23,24,25,26,28,29, 30,31,33}	90.30	92.67	72.95	88.96	51.99	76.13
V-shaped	V1	12	{1,3,7,12,21,22,23,24,28,29,31,33}	90.85	91.70	78.44	90.31	59.36	86.92
	V2	18	{1,3,4,7,8,10,12,15,19,20,21,22,24,25,26, 28,31,33}	90.38	92.24	81.79	89.66	53.40	76.13
	V3	12	{4,12,14,20,21,23,25,26,28,30,31,33}	90.99	92.18	75.37	87.39	48.85	76.13
	V4	9	{1,3,12,13,21,28,29,31,33}	90.42	91.19	84.69	90.11	58.17	87.38
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16, 17,18,19,20, 21,22,23,24,25,26,27,28,29,30,31,32,33}	90.74	93.35	79.03	90.18	51.96	76.13
Optimal subset	V4	9	{1,3,12,13,21,28,29,31,33}	90.42	91.19	84.69	90.11	58.71	87.38

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 11. Recognition rate of optimal feature subset with six classifiers in the case of six groups.

Activity	Optimal subsets	Recognition rate (%)					
		J48	RF	KNN	MLP	NB	SVM
Sitting–lying	{1,3,12,13,15,20,21,22,30}	99.15	99.38	98.88	99.41	98.41	96.82
Walking–lying	{1,8,14,15,21,23,29}	99.65	99.61	99.37	99.65	99.20	98.38
Standing–walking	{3,6,8,9,10,12,13,16,17,18,20,21,23,26,27,28,31,32}	87.68	88.39	87.05	88.37	84.90	65.94
Standing–lying	{2,7,14,17,19,23,28,29,30,31}	99.57	99.62	98.04	99.46	96.56	97.86
Walking–sitting	{2,4,8,9,10,12,13,16,17,21,22,27,30,31,33}	96.58	97.23	95.71	96.63	90.57	62.23
Standing–sitting	{1,3,12,13,21,28,29,31,33}	90.42	91.19	84.69	90.11	58.71	87.38

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.
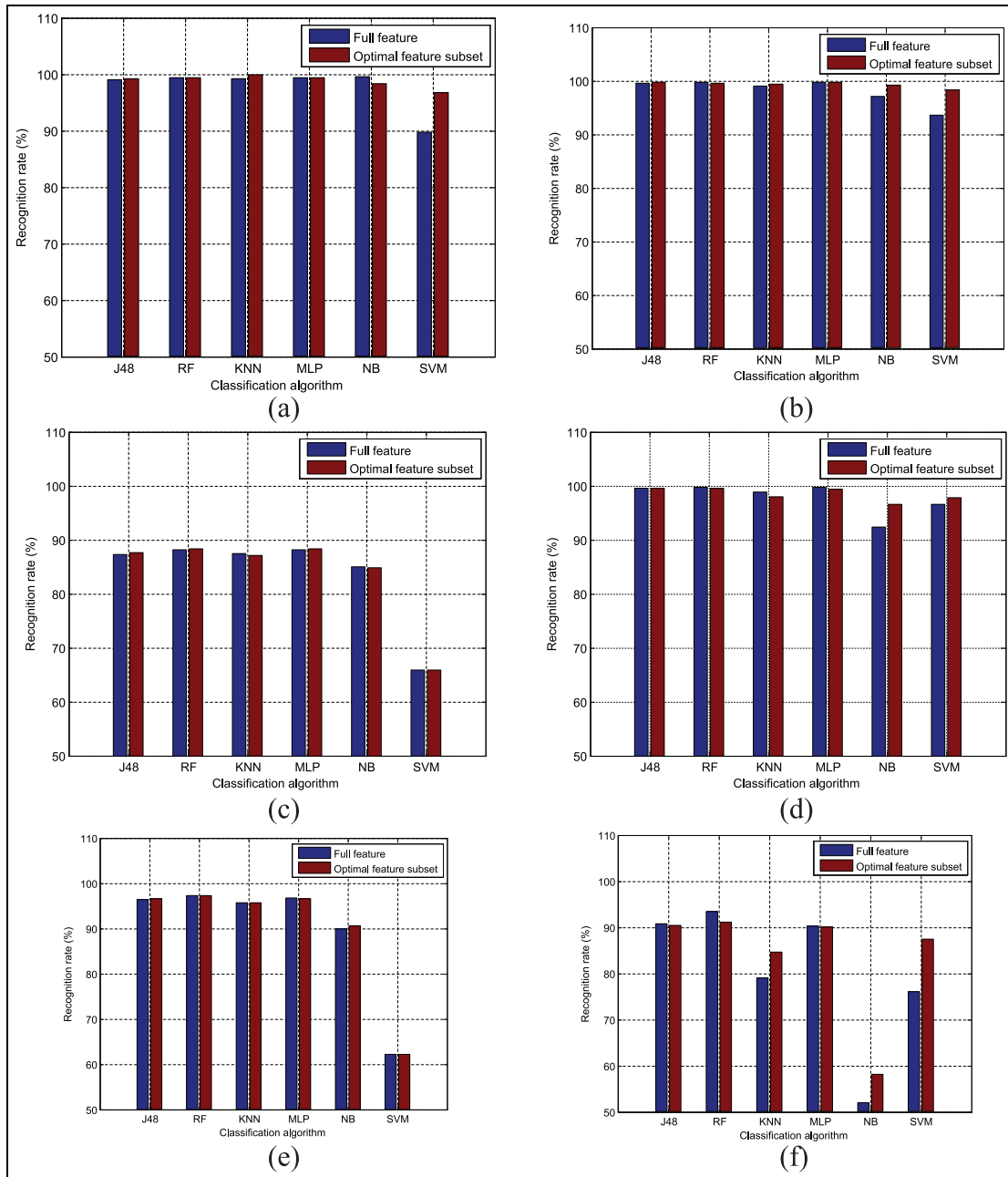

Figure 3. Recognition rate comparison of full feature and optimal feature subset for six activity groups: (a) sitting–lying, (b) walking–lying, (c) standing–walking, (d) standing–lying, (e) walking–sitting, and (f) standing–sitting, with a window size of 0.5 s and scheme 1 being conducted.

Table 12. Recognition rate and selected feature subset in the case of all activities (window size is 0.5 s, scheme 1).

Transfer function		Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	15	{1,2,3,9,11,12,13,15,18,19,20,21,22,29,31}	82.43	85.27	76.07	82.14	57.89	55.95
	S2	19	{1,2,4,7,10,11,12,15,19,20,21,22,24,25,26,27,30,31,33}	82.25	85.46	75.18	82.52	53.91	53.59
	S3	12	{9,13,14,19,21,22,23,24,28,29,30,31}	83.00	85.23	69.21	82.35	59.69	63.03
	S4	18	{2,4,9,12,13,14,19,20,22,23,25,26,27,28,29,30,31,33}	82.42	86.18	70.38	82.75	53.42	53.59
V-shaped	V1	10	{3,11,13,20,21,22,24,29,30,31}	82.19	85.08	68.66	82.18	59.90	62.94
	V2	9	{1,2,7,14,21,23,25,29,31}	83.04	85.20	79.52	83.18	56.10	58.06
	V3	15	{1,3,7,9,12,13,15,18,19,21,22,23,24,30,31}	81.66	84.32	72.23	81.90	58.45	57.60
	V4	13	{2,9,10,11,12,13,15,18,19,21,23,24,31}	82.15	84.54	74.43	81.82	57.44	56.01
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21, 22,23,24,25,26,27,28,29,30,31,32,33}	82.11	86.36	73.92	83.08	53.59	53.59
Optimal subset	V2	9	{1,2,7,14,21,23,25,29,31}	83.04	85.20	79.52	83.18	56.10	58.06

RF: random forest; KNN: k-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 13. Recognition rate and selected feature subset in the case of all activities (window size is 0.5 s, scheme 2).

Position of sensor	Scheme	Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
RKN	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	82.20	82.89	80.11	81.08	55.08	53.92
	Optimal subset	2	{8,10}	82.31	77.62	87.77	80.76	71.63	81.75
HIP	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	70.18	71.97	67.72	69.98	42.82	53.53
	Optimal subset	3	{3,7,8}	68.49	66.71	64.62	67.58	65.86	68.64
LUA	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	67.01	68.90	62.73	67.45	45.75	53.57
	Optimal subset	4	{1,3,4,9}	67.77	65.63	63.90	66.82	48.96	53.87

RF: random forest; KNN: k-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Features 1–10 represent the mean, variance, mean absolute deviation, root mean square, interquartile range, range, kurtosis, skewness, covariance, and energy at each position, respectively.

As can be seen from Table 13, the recognition effect of RKN is the best, followed by HIP, and that of LUA is worse. For all six classifiers, the recognition rates of J48, RF, and MLP are higher, KNN shows a medium recognition rate, while NB and SVM demonstrate the worst performance. The feature selection proposed in this article can effectively reduce the feature dimension, optimize the feature that contributes the most to the classification performance, and reduce the computational complexity as well as improve the recognition rate.

For the window size of 1 s, the total number of feature samples is 9078. The experimental results are shown in Table 14 and Figure 5 when the experiment was carried out according to scheme 1, and the experimental results of scheme 2 are shown in Table 15.

For the window size of 2 s, the total number of feature samples is 4536. The experimental results are shown in Table 16 and Figure 6 when the experiment was carried out according to scheme 1, and the experimental results of scheme 2 are shown in Table 17.

From Tables 12–17 and Figures 4–6, regardless of the size of the window, it follows that for each window size our proposed feature selection can effectively improve the recognition rate and reduce the computational complexity. Meanwhile, the recognition effect becomes better with the increase of window size. For a window size of 2 s, the recognition rate of RF can reach 94.49%. This is because the amount of data in each window increases with an increase of the window size. As such, the number of activity features embedded in the window also grows, leading to a higher recognition rate. In the cases with different window sizes, the recognition effect of sensors in the RKN position is the best and the most stable, followed by the HIP position, while the sensors placed in the LUA position show the worst effect. For the six classifiers, the recognition rates of J48, RF, and MLP are relatively high, KNN is of medium rate, while NB and SVM show the worst performance.

For the case with a window size of 2 s, the total number of feature samples is 4536 and the selection experiment is conducted with scheme 1. In the case of S-shaped and V-shaped conversion functions, the number of each dimension feature data belonging to the optimal feature subset is counted, as shown in Figure

Table 14. Recognition rate and selected feature subset under the condition of all activities (window size is 1 s, scheme 1).

Transfer function		Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	14	{1,2,10,12,13,14,20,21,22,23,24,28,30,31}	87.23	89.68	81.75	87.29	68.62	61.30
	S2	13	{2,3,10,12,13,19,20,21,22,23,24,30,31}	87.15	89.12	79.75	86.66	67.93	60.65
	S3	12	{2,13,14,19,20,21,22,23,28,29,30,31}	87.90	90.46	77.49	88.32	67.42	62.52
	S4	12	{11,12,13,19,20,21,22,23,28,29,30,31}	87.78	90.24	76.19	87.30	82.48	61.30
V-shaped	V1	11	{1,10,13,14,20,22,23,28,29,30,31}	87.39	89.64	80.94	88.33	69.05	64.34
	V2	11	{2,3,12,13,19,20,21,22,24,28,31}	87.23	89.31	78.59	86.69	67.49	60.05
	V3	12	{2,10,12,13,14,19,20,22,24,28,29,31}	88.10	90.28	81.08	88.58	67.54	61.01
	V4	13	{2,3,11,12,13,15,19,20,22,26,28,30,31}	87.41	90.16	78.01	86.44	65.21	53.88
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18, 19,20,21,22,23,24,25,26,27,28,29,30,31,32,33}	87.13	91.08	82.57	88.94	61.73	53.60
Optimal subset	V1	11	{1,10,13,14,20,22,23,28,29,30,31}	87.39	89.64	80.94	88.33	69.03	64.34

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 15. Recognition rate and selected feature subset under the condition of all activities (window size is 1 s, scheme 2).

Position of sensor	Scheme	Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
RKN	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	86.89	88.19	85.89	86.33	64.75	53.63
	Optimal subset	3	{5,8,10}	87.45	87.04	85.95	85.05	71.48	81.84
HIP	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	75.14	77.58	73.45	74.17	51.36	53.61
	Optimal subset	2	{3,8}	71.44	67.72	66.91	71.49	68.22	71.41
LUA	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	70.09	72.00	67.14	70.16	52.61	53.59
	Optimal subset	4	{1,4,5,9}	70.90	69.75	66.75	69.92	55.92	53.58

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

Table 16. Recognition rate and selected feature subset under the condition of all activities (window size is 2 s, scheme 1).

Transfer function		Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
S-shaped	S1	9	{3,7,13,14,22,23,24,28,31}	91.78	93.03	83.84	89.79	77.20	64.32
	S2	7	{3,7,13,22,23,30,31}	91.60	92.81	84.01	89.57	77.86	64.65
	S3	12	{3,7,9,12,13,20,21,22,24,28,29,31}	92.70	93.36	85.56	88.91	76.58	60.46
	S4	12	{2,12,13,15,21,22,23,24,28,29,30,31}	92.68	93.78	84.15	90.58	77.27	63.44
V-shaped	V1	15	{1,3,10,11,12,13,14,18,19,21,23,24,29,30,31}	93.50	94.18	88.05	92.50	79.05	56.10
	V2	11	{1,3,7,10,12,13,14,22,23,24,31}	92.48	92.92	89.92	90.52	78.50	67.73
	V3	11	{1,2,4,11,14,21,22,23,24,30,31}	92.72	93.47	87.32	91.73	78.77	55.70
	V4	10	{1,3,7,10,12,13,14,22,29,31}	92.94	93.80	90.94	91.42	78.57	63.84
No feature selection		33	{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18, 19,20,21,22,23,24,25,26,27,28,29,30,31,32,33}	93.34	94.49	88.49	92.79	72.24	53.61
Optimal subset	V1	15	{1,10,13,14,20,22,23,28,29,30,31}	93.50	94.18	88.05	92.50	79.05	56.10

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.

7. It can be seen from Figure 7 that the number of the 3rd, 13th, 22nd, 23rd, 24th, and 31st dimensions of the optimal feature subsets is the largest, but the final optimal feature subsets {1,3,10,11,12,13,14,18,19,21,23,24, 29,30,31} do not incorporate all of these higher-frequency features (that is, the optimal feature subset is not a simple combination of features with high frequency). The optimal feature subset does not necessarily include features with high frequency, since the feature subset selected by features of high frequency only may not have the best classification effect. On the contrary, features with lower frequency are also likely
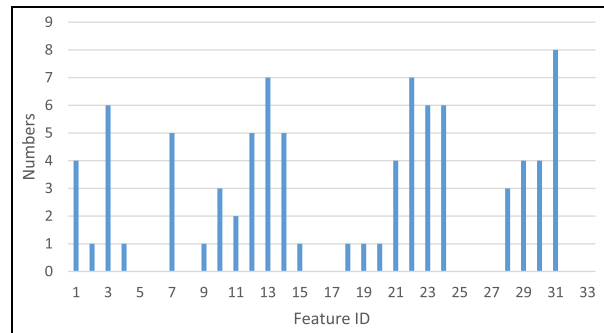
Table 17. Recognition rate and selected feature subset under the condition of all activities (window size is 2 s, scheme 2).

Position of sensor	Scheme	Dimension	Feature combination	Recognition rate (%)					
				J48	RF	KNN	MLP	NB	SVM
RKN	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	92.22	92.77	90.74	90.87	75.70	53.72
	Optimal subset	2	{3,4}	92.08	90.39	91.09	89.31	81.79	88.03
HIP	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	81.81	83.46	80.57	79.93	62.51	53.61
	Optimal subset	5	{3,5,7,8,10}	81.35	81.41	81.26	80.15	68.11	69.24
LUA	No feature selection	10	{1,2,3,4,5,6,7,8,9,10}	75.22	77.07	71.29	74.86	61.79	53.58
	Optimal subset	5	{1,5,7,8,10}	75.99	76.89	73.41	73.23	69.42	54.36

RF: random forest; KNN: *k*-nearest neighbor; MLP: multilayer perceptron; NB: naïve Bayesian; SVM: support vector machine.
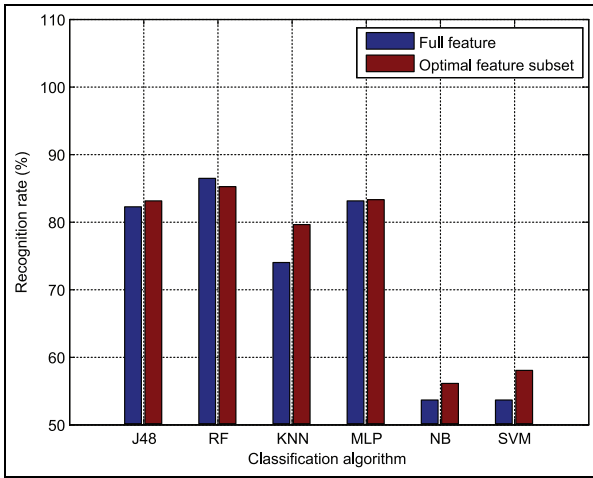
Figure 4. Comparison of recognition rates in the case of all activities (window size is 0.5 s, scheme 1).
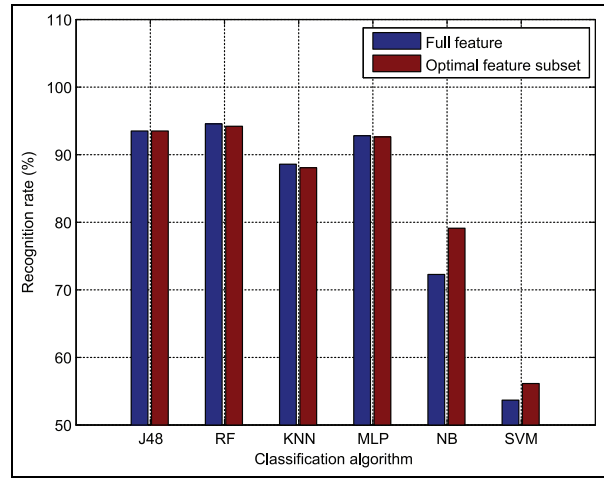
Figure 6. Comparison of recognition rates under the condition of all activities (window size is 2 s, scheme 1).
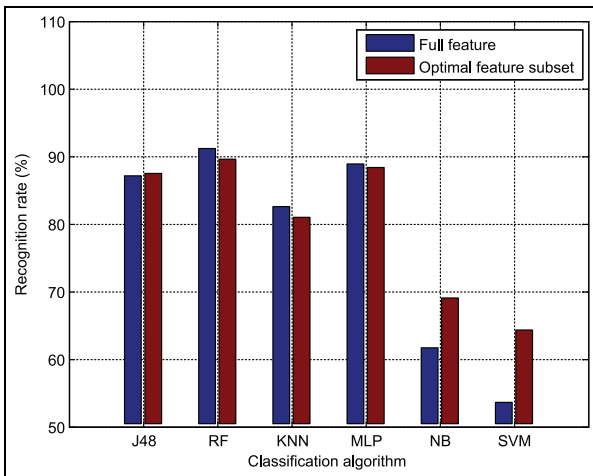
Figure 5. Comparison of recognition rates under the condition of all activities (window size is 1 s, scheme 1).

Figure 7. Number of each feature in the case of S-shaped and V-shaped transfer functions.

to generate stronger classification effect through the interaction of features, such as the 18th and 19th dimensions.

For the case with a window size of 2 s, the total number of feature samples is 4536 and the selection experiment is conducted in scheme 2. Figure 8 illustrates the convergence curves in the case of S-shaped and V-shaped transfer functions of three different position data, with four activities treated as a group. As can be seen from the figure, the eight types of transfer
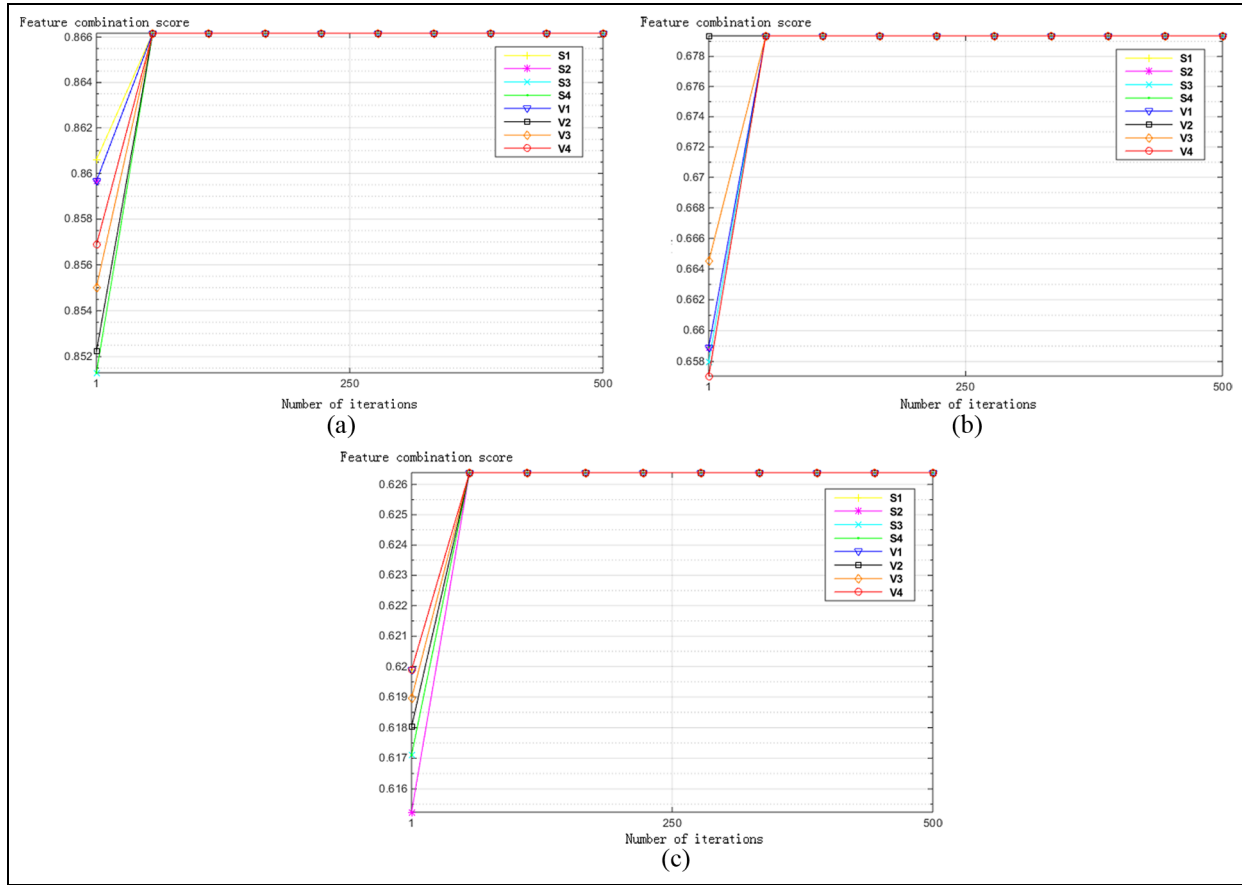
Figure 8. Convergence curves under the condition of S-shaped and V-shaped transfer functions: (a) RKN, (b) HIP, and (c) LUA.

functions converge rapidly under the three position conditions, which shows that the preferred method in this article can select the optimal subset of features that meet the requirements in a relatively short time and, to a certain extent, improves the real-time activity recognition system.

Conclusion

In this article, the feature selection problem in activity recognition is discussed first and then the BPSO is introduced briefly. Based on the criteria of classification accuracy, a correlation-based BPSO method for feature selection in human activity recognition is proposed. In the proposed algorithm, the correlation coefficients between the features are added to the BPSO as a feature correlation factor to determine the positions of particles. This method takes into account the particularity of the feature optimization problem and no longer uses the PSO algorithm as a black box. It makes the feature with more information more likely to be selected and improves the performance of the feature selection method based on PSO. The KNN classifier is then used as the fitness function in PSO to evaluate the performance of the feature subset, and the feature

combination with the highest KNN classifier recognition rate could be picked as the eigenvector. Finally, a large number of experiments were carried out for the proposed method. The results show that the proposed method can effectively improve the accuracy of activity recognition.

As we showed in this article, the proposed algorithm has a significant impact on activity recognition accuracy and should be taken into account. Considering the promising results obtained in this pilot study, we intend to work on novel feature selection methods based on BPSO while better dealing with data diversities. Another point that deserves to be further assessed is the optimization of runtime performance as it plays an important role in improving the recognition efficiency.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

1. Banos O, Damas M, Guillen A, et al. On the development of a real-time multi-sensor activity recognition system. In: *Proceedings of the 7th International Work-Conference IWAAL*, Puerto Varas, Chile, 1–4 December 2015. Berlin: Springer.
2. Liu Y, Nie L, Liu L, et al. From action to activity: sensor-based activity recognition. *Neurocomputing* 2016; 181: 108–115.
3. Reyes-Ortiz J-L, Oneto L, Sama A, et al. Transition-aware human activity recognition using smartphones. *Neurocomputing* 2016; 171: 754–767.
4. Bahrepour M, Meratnia N, Taghikhaki Z, et al. Sensor fusion-based activity recognition for Parkinson patients. In: Thomas C (ed.) *Sensor fusion—foundation and applications*. London: InTech, 2017, pp.171–190.
5. Wang L, Gu T, Tao X, et al. Sensor-based human activity recognition in a multi-user scenario. In: *Proceedings of the European conference on ambient intelligence*, Salzburg, 18–21 November 2009, pp.78–87. Berlin; Heidelberg: Springer.
6. Abdelniser M, Abdulbaset A and Ramahi OM. Reducing sweeping frequencies in microwave NDT employing machine learning feature selection. *Sensors* 2016; 16(4): 559.
7. Liu H, Yu L and Yu L. Toward integrating feature selection algorithm for classification and clustering. *IEEE T Knowl Data En* 2005; 17(4): 491–502.
8. González S, Sedano J, Villar JR, et al. Features and models for human activity recognition. *Neurocomputing* 2015; 167: 52–60.
9. Dash M and Liu H. Feature selection for classification. *Intell Data Anal* 1997; 1(3): 131–156.
10. Cerrada M, Sánchez RV, Cabrera D, et al. Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal. *Sensors* 2015; 15(9): 23903–23926.
11. Karegowda AG, Manjunath AS and Jayaram MA. Comparative study of attribute selection using gain ratio and correlation based feature selection, 2010, https://pdfs.semanticscholar.org/3555/1bc9ec8b6ee3c97c524f9c9ceee798c2026e.pdf
12. Theodoridis S and Koutroumbas K. *Pattern recognition*. Oxford: Academic Press, 2008.
13. Raileanu LE and Stoffel K. Theoretical comparison between the Gini index and information gain criteria. *Ann Math Artif Intell* 2004; 41: 77–93.
14. Xiaofei H, Cai D and Niyogi P. Laplacian score for feature selection. *Adv Neural Inf Pr Syst* 2005; 18: 507–514.
15. Gu Q, Li Z and Han J. Generalized fisher score for feature selection. In: *Proceedings of the international conference on uncertainty in artificial intelligence*, Barcelona, Spain, 14–17 June 2011. New York: ACM.
16. Peng H, Long F and Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE T Pattern Anal Mach Intell* 2005; 27: 1226–1238.
17. Lai C, Reinders MJT and Wessels L. Random subspace method for multivariate feature selection. *Pattern Recogn Lett* 2006; 27: 1067–1076.
18. Ferreira AJ and Figueiredo MAT. An unsupervised approach to feature discretization and selection. *Pattern Recogn* 2012; 45: 3048–3060.
19. Tabakhi S, Moradi P and Akhlaghian F. An unsupervised feature selection algorithm based on ant colony optimization. *Eng Appl Artif Intell* 2014; 32: 112–123.
20. Tabakhi S and Moradi P. Relevance–redundancy feature selection based on ant colony optimization. *Pattern Recogn* 2015; 48: 2798–2811.
21. Moradi P and Rostami M. A graph theoretic approach for unsupervised feature selection. *Eng Appl Artif Intell* 2015; 44: 33–45.
22. Moradi P and Rostami M. Integration of graph clustering with ant colony optimization for feature selection. *Knowl Based Syst* 2015; 84: 144–161.
23. Gheyas IA and Smith LS. Feature subset selection in large dimensionality domains. *Pattern Recogn* 2010; 43: 5–13.
24. Deepa O and Senthilkumar A. Swarm intelligence from natural to artificial systems. *Ant Colon Optim* 2016; 8(1): 9–17.
25. Wang X, Yang J, Teng X, et al. Feature selection based on rough sets and particle swarm optimization. *Pattern Recogn Lett* 2007; 28: 459–471.
26. Sikora R and Piramuthu S. Framework for efficient feature selection in genetic algorithm based data mining. *Eur J Oper Res* 2007; 180: 723–737.
27. Chtioui Y, Bertrand D and Barba D. Feature selection by a genetic algorithm: application to seed discrimination by artificial vision. *J Sci Food Agric* 2015; 76(1): 77–86.
28. Farmer ME, Bapna S and Jain AK. Large scale feature selection using modified random mutation hill climbing. In: *Proceedings of the 17th international conference on pattern recognition*, Cambridge, 26 August 2004, pp.287–290. New York: IEEE.
29. Skalak DB. Prototype and feature selection by sampling and random mutation hill climbing algorithms. In: *Proceedings of the 11th international conference on machine learning*, New Brunswick, NJ, 10–13 July 1994, pp.293–301. New York: ACM.
30. Meiri R and Zahavi J. Using simulated annealing to optimize the feature selection problem in marketing applications. *Eur J Oper Res* 2006; 171: 842–858.
31. Forsati R, Moayedikia A and Keikha A. A novel approach for feature selection based on the bee colony optimization. *Int J Comput Appl* 2012; 43(8): 13–16.
32. Schiezaro M and Pedrini H. Data feature selection based on artificial bee colony algorithm. *EURASIP* 2013; 2013: 47.
33. UncuÖ and Türkşen IB. A novel feature selection approach: combining feature wrappers and filters. *Inform Sci* 2007; 177(2): 449–466.

34. Hu Q, Pedrycz W, Yu D, et al. Selecting discrete and continuous features based on neighborhood decision error minimization. *IEEE T Syst Man Cy B* 2010; 40(1): 137–150.

35. Xie J and Wang C. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. *Expert Syst Appl* 2011; 38(5): 5809–5815.

36. Karagiannaki K, Panousopoulou A and Tsakalides P. An online feature selection architecture for human activity recognition. In: *Proceedings of the international conference on acoustics speech and signal processing*, New Orleans, LA, 5–9 March 2017. New York: IEEE.

37. Chernbumroong S, Cang S and Yu H. Maximum relevancy maximum complementary feature selection for multi-sensor activity recognition. *Expert Syst Appl* 2015; 42(1): 573–583.

38. Fang H, He L, Si H, et al. Human activity recognition based on feature selection in smart home using back-propagation algorithm. *ISA Trans* 2014; 53(5): 1629–1638.

39. Oukrich N, Maach A, Sabri E, et al. Activity recognition using back-propagation algorithm and minimum redundancy feature selection method. In: *Proceedings of the 2016 4th IEEE international colloquium on information science and technology (CIST)*, Tangier, Morocco, 24–26 October 2016, pp.818–823. New York: IEEE.

40. Suto J, Oniga S and Sitar PP. Comparison of wrapper and filter feature selection algorithms on human activity recognition. In: *Proceedings of the 2016 6th international conference on computers communications and control (ICCCC)*, Oradea, 10–14 May 2016, pp.124–129. New York: IEEE.

41. Uddin MT and Uddiny MA. A guided random forest based feature selection approach for activity recognition. In: *Proceedings of the 2015 international conference on electrical engineering and information communication technology (ICEEICT)*, Dhaka, 21–23 May 2015, pp.1–6. New York: IEEE.

42. Wang Z and Huo Y. A multi-attribute fusion acceleration feature selection algorithm for activity recognition on smart phones. In: *Proceedings of the 2014 international conference on information science, electronics and electrical engineering (ISEEE)*, Sapporo, 26–28 April 2014, pp.145–148. New York: IEEE.

43. Mazaar H, Emary E and Onsi H. Ensemble based-feature selection on human activity recognition. In: *Proceedings of the international conference on informatics and systems*, Giza, Egypt, 9–11 May 2016, pp.81–87. New York: ACM.

44. González S, Sedano J, Villar JR, et al. Features and models for human activity recognition. *Neurocomputing* 2015; 167(C): 52–60.

45. Nunes UM, Faria DR and Peixoto P. A human activity recognition framework using max-min features and key poses with differential evolution random forests classifier. *Pattern Recogn Lett* 2017; 99: 21–31.

46. Xian X, Xianling L, Wang H, et al. Accelerometer data feature selection for activity recognition based on GA optimization. *Comput Eng Appl* 2016; 52: 139–143.

47. Kennedy J. *Particle swarm optimization, Encyclopedia of Machine Learning*. Berlin: Springer, 2011, pp.760–766.

48. Chen Q, Chen Y and Jiang W. Genetic particle swarm optimization–based feature selection for very-high-resolution remotely sensed imagery object change detection. *Sensors* 2016; 16(8): 1204.

49. Kennedy J and Eberhart RC. A discrete binary version of the particle swarm algorithm. In: *Proceedings of the IEEE international conference on 1997 computational cybernetics and simulation systems, man, and cybernetics*, Orlando, 12–15 October 1997, pp.4104–4108. New York: IEEE.

50. Haixiang G, Yijing L, Yanan L, et al. BPSO-Adaboost-KNN ensemble learning algorithm for multi-class imbalanced data classification. *Eng Appl Artif Intell* 2016; 49: 176–193.

51. Mirjalili S and Lewis A. S-shaped versus V-shaped transfer functions for binary particle swarm optimization. *Swarm Evol Comput* 2013; 9: 1–14.

52. Roggen D, Calatroni A, Rossi M, et al. Collecting complex activity datasets in highly rich networked sensor environments. In: *Proceedings of the 2010 seventh international conference on networked sensing systems (INSS)*, Kassel, 15–18 June 2010, pp.233–240. New York: IEEE.

53. Chavarriaga R, Sagha H, Calatroni A, et al. The Opportunity challenge: a benchmark database for on-body sensor-based activity recognition. *Pattern Recogn Lett* 2013; 34(15): 2033–2042.

54. Noor MHM, Salcic Z and Wang IK. Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer. *Pervasive Mob Comput* 2016; 38: 102–107.

55. Qin N, Patterson T, Cleland I, et al. Dynamic detection of window starting positions and its implementation within an activity recognition framework. *J Biomed Inform* 2016; 62(C): 171–180.

56. Banos O, Galvez JM, Damas M, et al. Window size impact in human activity recognition. *Sensors* 2014; 14(4): 6474–6499.