Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers

Jhun-Ying Yang, Jeen-Shing Wang *, Yen-Ping Chen

Department of Electrical Engineering, National Cheng Kung University, 1 University Road, Tainan 701, Taiwan

Abstract

This paper presents a systematic design approach for constructing neural classifiers that are capable of classifying human activities using a triaxial accelerometer. The philosophy of our design approach is to apply a divide-and-conquer strategy that separates dynamic activities from static activities preliminarily and recognizes these two different types of activities separately. Since multilayer neural networks can generate complex discriminating surfaces for recognition problems, we adopt neural networks as the classifiers for activity recognition. An effective feature subset selection approach has been developed to determine significant feature subsets and compact classifier structures with satisfactory accuracy. Experimental results have successfully validated the effectiveness of the proposed recognition scheme.

1. Introduction

Human–computer interaction (HCI) is a notable discipline that bridges the gap between users and computer systems, and is increasingly being recognized as an indispensable component of daily life. One of the key techniques in HCI is pattern recognition. Users’ intentions can be recognized using recognition techniques without using the traditional input devices of computer systems. Among various pattern recognition issues, activity recognition is a new technique which can recognize human activities or gestures via computer systems, and the signal for recognition can be obtained from different kinds of detectors such as electromyography (EMG), audio sensors, image sensors, and accelerometers. Due to the rapid development of technology and the omnipresence of reasonable low-cost high-performance personal computers, research on human activity recognition has grown up and activity recognition is being applied in many applications including biomedical engineering, medical nursing, and interactive entertainment (Choi et al., 2005; Delsey et al., 2005; Najafi et al., 2003; Song and Wang, 2005). Among the aforementioned sensors for activity recognition, accelerometers can return a real-time measurement of acceleration along the x-, y- or z-axis to be used as a human motion detector. Due to advanced miniaturization techniques, the accelerometer can be embedded within a wearable device and can send data wirelessly to a mobile computing device. This greatly reduces the users’ awareness and possible discomfort during the process of data collection and recognition.

Decision tree classification methods have been successfully used for recognition problems as well as the activity recognition from acceleration data (Bao and Intille, 2004; Mathie et al., 2004; Karantonis et al., 2006; Solà i Carós et al., 2005). To name a few, Bao and Intille (2004) used five biaxial accelerometers placed on 20 subjects (in five positions: arm, wrist, hip, angle, and thigh) to recognize 20 daily human activities. They designed a data collection process annotated by the subjects themselves under semi-naturalistic conditions and compared the performances of four recognition methods: the decision table, instance-based learning (IBL), C4.5 decision tree, and Naïve Bayes classifiers. C4.5 decision trees obtained the best performance with an overall recognition accuracy rate of 84%. The decision tree method usually separates static activities from dynamic activities first, and a more detailed subclassification is made at the following hierarchical structures (Karantonis et al., 2006; Mathie et al., 2004; Solà i Carós et al., 2005). The advantages of decision trees are that they are simple, apparent, and fast in reasoning. However, traditional decision trees need to hold considerable data in each of their non-terminal roots, which increases the required memory space and computation time.

Recently, neural network techniques have provided an alternative approach to pattern recognition due to its learning capability of separating non-linearly separable classes. Neural networks can
autonomously learn the complex mappings and extract a non-linear combination of features. The weights of the networks create the decision boundaries in the feature space, and the resulting discriminating surfaces can be very complex (Safavian and Landgrebe, 1991). In general, placing more accelerometers on different body positions can recognize human movements more accurately (Bao and Intille, 2004). But this will cost more, be less comfortable, and require a higher dimension and more complex feature set. So far, only a few researchers have devoted their efforts to the investigation of recognizing uncomplicated activities via one accelerometer (Karantonis et al., 2006; Mathie et al., 2004; Solà i Carós et al., 2005).

In this paper, we adopted multilayer feedforward neural networks (FNNs) as activity classifiers and proposed an effective activity recognition method using acceleration data. The proposed approach can recognize more complicated daily activities via the use of one triaxial accelerometer on a wearable board mounted on the dominant wrist of a subject to acquire the acceleration data of his/her activity. We collected acceleration data for a set of eight common domestic activities in a controlled laboratory environment: standing, sitting, walking, running, vacuuming, scrubbing, brushing teeth, and working at a computer. The philosophy of our recognition approach is to apply a divide-and-conquer strategy that separates dynamic activities (e.g. walking, running, etc.) from static activities (e.g. standing and sitting) preliminarily and recognizes these two different types of activities separately. First, we used a neural classifier as a pre-classifier and adopted the constant parameter (Karantonis et al., 2006) to distinguish static activities from dynamic activities. The development of an effective feature subset selection (FSS) approach based on the common principal component analysis (CPCA) proposed in (Krzanowski, 1979) has been conducted to reduce the dimension of the feature sets for static and dynamic activities, respectively. This approach can determine the significant feature subsets and retain the characteristics of the data distribution in feature spaces. The selected features were then used to construct neural classifiers for dynamic and static activities. Our experiments on recognizing eight daily activities from seven subjects have successfully validated the effectiveness of the proposed recognition scheme in constructing efficient classifiers with satisfactory accuracy.

The remaining sections of this paper are organized as follows. Section 2 introduces the structure of the neural classifier. The detailed information about the proposed recognition method, including data pre-processing, feature extraction, and feature subset selection, is presented in Section 3. Section 4 provides the experimental design to validate the effectiveness of the proposed classifiers and discussions on the recognition results. Conclusions are given in the last section.

2. Structure of neural classifier

In this paper, three neural classifiers, including a pre-classifier, a static classifier, and a dynamic classifier, are constructed to recognize daily activities from acceleration data. The structure of the neural classifier, shown in Fig. 1, consists of an input layer, a hidden layer, and an output layer. \( u = [u_1, u_2, \ldots, u_r]^T \) and \( y = [y_1, y_2, \ldots, y_h]^T \) are the input and output vectors, respectively, where \( r \) represents the number of elements in the input feature set and \( h \) is the number of classes. Log sigmoid functions are selected as the activation functions \( f \) in the hidden and output neurons. In general, the backpropagation learning algorithm (a gradient descent optimization method) is used to train the FNN. However, it is known that the gradient descent learning method is subject to slow convergence and local minima. Some remedies have been provided by Hannan and Bishop (1997) and Pearlmutter (1995) for accelerating the convergence and for achieving optimal solutions.

The resilient backpropagation (RPROP) learning algorithm (Riedmiller and Braun, 1993) is one of the best solutions for neural network training. Research studies have shown that RPROP can remedy the drawbacks of the gradient descent (Igel and Hasken, 2003). The basic principle of RPROP is to ignore the magnitude of the gradient and only take the sign of the derivative into consideration for indicating the direction of the weight update. To accelerate the construction of classifiers, the RPROP algorithm has been adopted to train our neural classifiers. The detailed information about the RPROP can be found in (Riedmiller and Braun, 1993).

3. Activity recognition strategy

To perform activity recognition robustly, reliably, and accurately, we have developed a neural classifier construction scheme. The proposed scheme consists of two phases: pre-classifier construction phase and static/dynamic classifier construction phase, and their corresponding block diagrams are shown in Figs. 2 and 3, respectively. The objective of the pre-classifier construction phase is to identify static activities and dynamic activities. First, we filter the acceleration data to obtain human body acceleration (BA). Then the features extracted from the BA are used to train the pre-classifier. Upon completion of the pre-classifier construction, we are able to distinguish dynamic activities from static activities. In the static/dynamic classifier construction phase, we extract various features from the original acceleration data into a feature set. In order to reduce the dimension of the feature set, we utilize a feature subset selection (FSS) approach to select significant features for static and dynamic activity, respectively. The corresponding selected feature sets are fed into and used to train the static/dynamic classifiers.

3.1. Pre-classifier construction

3.1.1. Data pre-processing

In general, the acceleration data acquired from the triaxial accelerometer on the human body can be decomposed into two components consisting of gravitational acceleration (GA) and body acceleration (BA). The BA component caused by body movement is able to distinguish static activities from dynamic activities.
We utilize a high pass filter to obtain the BA component by removing the GA from the raw acceleration signals.

Due to the long and continuous sequence of the acceleration data, it is difficult to analyze and recognize activities from this huge data sequence without any manipulation. Consequently, cutting the acceleration sequences into many overlapping windows (segments) of the same length is a preferable way. Fig. 4 illustrates the tri-axial acceleration data with overlapping windows. Features will be extracted from the filtered acceleration data with windows having an identical size. A target tag vector of a window which belongs to static activity will be assigned \([1,0]^T\); otherwise, it will be assigned \([0,1]^T\).

### 3.1.2. Feature extraction

Extracting features from a window is a fairly effective way to preserve class separability and can represent the characteristics of different activity signals in each window. We introduce the following two features extracted from the BA component to identify static and dynamic activities. (1) Signal magnitude area (SMA) (Bouten et al., 1997; Karantonis et al., 2006): the SMA is equal to the sum of acceleration magnitude summations over three axes of each window normalized by the window length. The discrete form of the SMA can be given by

\[
SMA = \frac{1}{w} \left( \frac{1}{w} \sum_{i=1}^{w} |x_i| + \frac{1}{w} \sum_{i=1}^{w} |y_i| + \frac{1}{w} \sum_{i=1}^{w} |z_i| \right),
\]

where \(w\) is the window length; \(x_i, y_i\), and \(z_i\) represent \(i\)th BA components of the \(x\)-, \(y\)-, and \(z\)-axis samples in a window, respectively; (2) average energy (AE): energy is calculated as the sum of the squared discrete FFT component magnitudes of the signal in a window (Wang et al., 2005) and we adopt the average of the energy over three axes as the second feature.
Note that the SMA and AE are scalars. Obviously, the subject is in a resting state if the above features are quite small; otherwise, the subject is in an active state. Therefore, we can utilize the error vectors between the target tag and the output of the pre-classifier fed by the two features to fine-tune the parameters of the pre-classifier. The construction process of the pre-classifier is shown in Fig. 2.

3.2. Static/dynamic classifier construction

3.2.1. Data pre-processing

Before constructing the static/dynamic classifiers, we window the raw acceleration data without filtering and annotate each window using a target tag vector. The ith element of the target tag vector is “1” and the others are “0” as the corresponding window belongs to the ith class.

3.2.2. Feature extraction

To recognize the different classes of static and dynamic activities, we introduce eight features, consisting of mean, correlation between axes, energy, interquartile range, mean absolute deviation, root mean square, standard deviation, and variance, which are usually extracted from the triaxial acceleration data (Maurer et al., 2005; Ravi et al., 2005; Wang et al., 2005) and explicated as follows: (1) mean: the mean value of the acceleration data over a window is the DC component of the signal; (2) correlation between axes: correlation is especially useful for discriminating between activities that involve translation in just one dimension. For instance, walking and climbing stairs may have a similar periodicity and magnitude of the acceleration signal on the limbs, but it usually involves translation in one dimension when walking, whereas more than one dimension is involved in climbing stairs (Bao and Intille, 2004); (3) energy: due to the visibility of the periodicity of the acceleration data in the frequency domain, the energy feature is to capture data periodicity and can be used to discriminate sedentary activities from moderate and vigorous ones; (4) interquartile range: when the mean values of different classes are similar, the interquartile range represents the dispersion of the data and avoids the effect on range caused by extreme values in the data. And the following four features are different approaches but have similar purposes that measure the magnitude of a varying quantity in the acceleration data; (5) mean absolute deviation; (6) root mean square; (7) standard deviation; and (8) variance.

Since the triaxial accelerometer collects signals from three axes, x-, y-, and z-axis, a total of 24 (3 x 8) features are calculated from a window of the acceleration data.

3.2.3. Feature subset selection

The input variables of the proposed neural classifier are the feature set extracted from the acceleration data. In general, if the dimension of a feature set is too high, it may result in the following problems for classification and neural network training: (1) some features are irrelevant or redundant and do not provide supportive information to significantly improve the recognition accuracy; (2) the computational speed may be slow and the training of neural networks in the high dimension space is difficult. In the static/dynamic classifier construction, we take 24 features into consideration. However, this may be a considerable quantity of features for training the neural classifiers. Thus, the dimensionality reduction technique for feature extraction is taken into account in the pre-processing step. The principal component analysis (PCA) is one of the well-known methods for multivariate statistical analysis and can transform the original features into a lower dimensional space. However, an undesirable constraint of the PCA is that when applying the PCA, all data groups must be arranged into one group. To remove such a constraint, the common principal component analysis (CPCA) (Krzakowski, 1979; Flury, 1984), generalized from the PCA, has been proposed for multivariate analysis in multi-classes that contain the same variables. Recently, Yoon et al. (2005) proposed an unsupervised feature subset selection method based on the CPCA and applied the k-means clustering technique to reduce dimensionality by choosing a subset of the original features. Here, we propose a supervised feature subset selection method modified from (Yoon et al., 2005) to reduce the dimension of the feature set.

We denote the features collected from all subjects for one activity as a feature data set (FDS). Assume that we have N FDSs (N activities). Each FDS can be represented as an \(m \times n\) matrix, where \(m\) is the total number of windows for all the subjects and \(n\) is the number of features. Krzakowski (1979) showed that the common principal component (CPC) loadings provide the successively components that agree most closely with every principle component (PC) of the FDSs and the first CPC indicates the common direction along which each FDS item has its maximum variance (Yoon et al., 2005). That is, the CPC loadings retain the characteristics of the data distribution of the FDSs (activities) simultaneously. For example, Fig. 5 shows the distribution of two classes (“o” and “+”), where the dashed line and the dotted line represent the first PCs of each class, respectively. The solid line is the first CPC which bisects the angle between the two PCs and shows that the two classes have their maximum variance along the common direction of the CPC. In the CPC loadings, the ith row vector represents the projection of the ith feature of the FDSs to the lower dimensional common subspace. According to the investigation of Cohen et al. (2000), if features are highly correlated, the corresponding row vectors in the loadings are similar. Therefore, a clustering technique is useful to group similar loadings for the feature subset selection. We adopt the support vector clustering (SVC) algorithm, first proposed by Ben-Hur et al. (2001), to cluster the row vectors of the CPC loadings to indicate which features have similar contributions to each of the CPC loadings. In the SVC algorithm, data points are mapped from the data space to a high-dimensional feature space via a Gaussian kernel transformation. The idea of the SVC is to search the smallest sphere in the feature space that encloses the images of the data. The contours of clusters in the data

![Fig. 5. First common principle component of two classes.](image-url)
space can be generated by finding the support vectors on the sphere surface in the feature space. These contours are interpreted as cluster boundaries. Now we present the proposed FSS algorithm as follows:

**Step 1**: Apply the PCA on each FDS item to obtain the loadings. Let the loading matrix of the kth FDS item be denoted as \( L_k \), \( k = 1, \ldots, N \). The \((i,j)\)th element in \( L_k \) represents the loading value of the ith feature with respect to the jth PC for the kth FDS item.

**Step 2**: Although there are n PCs for each FDS, only the first \( p_k \) PCs can be sufficient to represent the kth FDS item. In general, \( p_k \) is determined based on the cumulative contribution of the PCs. We denote \( \lambda_i(k) \) as the ith singular value of the covariance matrix of the kth FDS item. The cumulative contribution of first \( p \) PCs is calculated by:

\[
\text{cumContribution}(k) = 100 \times \frac{\sum_{i=1}^{p} \lambda_i(k)}{\sum_{i=1}^{n} \lambda_i(k)}
\]

**Step 3**: According to Krzanowski (1979), the CPC loadings are defined by the eigenvectors of the matrix \( H = \sum_{k=1}^{N} U_k U_k^T \). The singular value decomposition (SVD) is applied to decompose \( H \), such that \( H = VSV^T \), where \( S \) is a \( n \times n \) diagonal matrix whose diagonal elements are the singular values of \( H \), and \( V \) is the corresponding eigenvectors (components) of \( H \). The first \( p \) columns of \( V \) are the CPC loadings denoted as \( V_p \).

**Step 4**: We take the row vectors of \( V_p \) as the data points and perform the SVC algorithm to find the number of clusters (\( K \)). Then, the most representative features can be obtained by selecting the corresponding points closest to the centroids of the \( K \) clusters and identifying their corresponding original features. For instance, there are a total of 24 features and the distribution of the first two components \( (p = 2) \) of \( H \) obtained by Step 3 is shown in Fig. 6. We can find that there are five clusters \( (K = 5) \) in the data distribution. Consequently, the five features whose row vectors closest to the centroid of each cluster can be utilized to represent all the features.

We perform the proposed FSS method for static and dynamic activities, respectively, to determine parsimonious static and dynamic feature subsets. Then, we feed the values of the corresponding feature subsets to the static/dynamic classifiers and use the target tags generated in the data pre-processing phase to train the neural classifiers. Fig. 3 illustrates the overall construction procedure of the static/dynamic classifiers.

### 3.3. Verification

Fig. 7 shows the block diagram of performing the trained neural classifiers to recognize activity from acceleration data. An unknown sequence of acceleration data is first filtered by an elliptical...
high pass filter to remove the effect of gravity. Next, we window the filtered sequence and extract the features, the SMA and the AE, into a feature vector. Two elements of the output vector of the trained pre-classifier fed by the feature vector will compete against each other, and we set the winner element at one and the lost element at zero to indicate the fundamental type of the unknown activity, i.e., \([1,0]^T\) represents static activities and \([0,1]^T\) represents dynamic activities. Upon completion of the above procedure, the window has been labeled static or dynamic. We re-window the unknown sequence without filtering and extract the corresponding selected feature subset vector according to the result of the proposed FSS method. Then, we feed the vector to the trained static or dynamic classifier and each element in the output vector competes with the others. Similarly, the winner element will be set to “1” and the lost elements to “0”. Consequently, the recognition result of the unknown sequence is decided after the above procedure. Note that the FSS method is not utilized since we have already decided what signification feature should be taken; thus, we do not extract all the features but just calculate the selected feature sets in the verification stage.

4. Experimental design and results

The acceleration data was collected using the MMA7260Q triaxial accelerometer, designed by Freescale Semiconductor, with a microcontroller (C8051F330) and a wireless transceiver (nRF2401) on a wearable board. The sensor module satisfies the fundamental requirements of the acceleration device: lightness, sensing, and wireless transmission. The accelerometer’s sensitivity is set from $-4.0 \text{ g}$ to $+4.0 \text{ g}$ and the output signal of the accelerometer is sampled at 100 Hz by a 10-bit ADC. Our recognition experiments were performed on a PC running Microsoft Windows XP operating system with a P4 Duo 2.4 GHz CPU and 2GB memory.

In our study, we focused on eight common domestic activities: standing, sitting, walking, running, vacuuming, scrubbing, brushing teeth, and working at a computer. We gathered acceleration data from seven normal, healthy subjects (4 females, 3 males; age $24.1 \pm 1.8$ years) in a controlled laboratory environment; that is, a clear definition of each activity is provided and specified. A single triaxial accelerometer module was mounted on the domi-

![Figure 8](image-url) The acceleration data over x-, y-, and z-axis collected from the first subject.

![Figure 9](image-url) Performance comparison with different numbers of features.

![Table 1](image-url) Recognition results of the cross-validation procedure for all the subjects

<table>
<thead>
<tr>
<th>Cross-validation</th>
<th>Number of selected features</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Static</td>
<td>Dynamic</td>
</tr>
<tr>
<td>First subject</td>
<td>2 (3)</td>
<td>9 (6)</td>
</tr>
<tr>
<td>Second subject</td>
<td>5 (4)</td>
<td>8 (5)</td>
</tr>
<tr>
<td>Third subject</td>
<td>6 (3)</td>
<td>8 (6)</td>
</tr>
<tr>
<td>Fourth subject</td>
<td>4 (3)</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Fifth subject</td>
<td>2 (3)</td>
<td>6 (5)</td>
</tr>
<tr>
<td>Sixth subject</td>
<td>5 (3)</td>
<td>7 (5)</td>
</tr>
<tr>
<td>Seventh subject</td>
<td>5 (3)</td>
<td>7 (5)</td>
</tr>
<tr>
<td>Average</td>
<td>95.24</td>
<td>87.18</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>3.76</td>
<td>7.37</td>
</tr>
</tbody>
</table>

(\(^T\)): Maximum of the number of the first $p_k$ principal components whose cumulative sum $>90\%$. 

![Graph](image-url)
Table 2  
Confusion matrix for all the subjects

<table>
<thead>
<tr>
<th>Classified as</th>
<th>Walking</th>
<th>Running</th>
<th>Scrubbing</th>
<th>Standing</th>
<th>Working at a PC</th>
<th>Vacuuming</th>
<th>Brushing teeth</th>
<th>Sitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walking</td>
<td>313</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Running</td>
<td>0</td>
<td>313</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Scrubbing</td>
<td>0</td>
<td>0</td>
<td>278</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>Standing</td>
<td>0</td>
<td>0</td>
<td>315</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Working at a PC</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>286</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Vacuuming</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>296</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Brushing teeth</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>288</td>
<td>0</td>
</tr>
<tr>
<td>Sitting</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>311</td>
<td></td>
</tr>
</tbody>
</table>

nart wrist, which is better for discriminating activities involving upper body movements (Bao and Intille, 2004). All the subjects were asked to perform the eight tasks for 2 min per activity. Fig. 8 shows the first 1500 acceleration data over x-, y- and z-axis of the eight activities collected from the first subject. Note that, since the sampling frequency is 100 Hz, the total number of the acceleration data for each activity of each subject is 12,000. Here, we utilized a leave-one-subject-out cross-validation method to validate the effectiveness of the proposed neural classifier construction scheme. In each repetition of the cross-validation process, one minute of the acceleration data collected from six subjects was used in the training procedure of the recognition scheme. Then, 2 min of the acceleration data from the subject who was left out of the training procedure were used to test the recognition performance. We repeat the procedure for all the subjects. The effectiveness of feature extraction on windows with 50% overlap has been validated in several studies (Bao and Intille, 2004; Wang et al., 2005). Here, we take the window size of 512 with 256 samples overlapping among consecutive windows. The 512-sample size enables fast computation of FFTs for calculating frequency-domain features. Each window represents 5.12 s at a sampling frequency of 100 Hz. After windowing, we can obtain 22 windows in one minute and 45 windows in 2 min for each activity of one subject. When we perform each repetition of the leave-one-subject-out training procedure, the total numbers of the training data and test data are $22 \times 8 \times 6$ and $45 \times 8$, respectively.

To evaluate the effectiveness of the proposed FSS method, we experimented on the leave-one-subject-out cross-validation procedure with the number of dynamic selected features from 2 to 24 and six static selected features. Note that we did not use the FSS method when the number of the selected features is 24. Fig. 9 illustrates the comparison of the average performance of different numbers of selected features. From Fig. 9, we found that the performances of the recognition scheme with extra features are pretty similar to the performance of the scheme with almost 1/3 quantity of features. Hence, we can confirm the effectiveness of the proposed FSS method.

The best average recognition accuracy is 95.24% in the cross-validation procedure. The numbers of neurons in the hidden layers are 3, 5 and 7 for the pre-classifier, static classifier and dynamic classifier, respectively, and the number of epochs is 500 for each training. Table 1 shows the recognition results of the cross-validation procedure for all the subjects with the corresponding numbers of the selected features. Note that, due to different training data used in each repetition of the cross-validation process, we consider higher accuracy and lower computation burden as a criterion for determining what number of selected features should be chosen. In addition, Table 1 also shows the performance comparison of the proposed neural classifiers with the k-nearest neighbor (KNN) method. From Table 1, we found that our neural classifiers outperform the KNN method.

To justify the necessity of the proposed feature subset selection (FSS), we have carefully analyzed the computational load, efficiency of the training procedure, and recognition accuracy. For the computational load of the training procedure, we evaluated the execution times of the classifier training with and without our FSS method. In the training procedure, the average execution time of the classifier training without the FSS step was 7.457 s while the execution time of the algorithm with the FSS step was 8.46 s. The FSS step requires additional 13.45% time consumption in performing the SVC. However, the average number of adjustable parameters for the subjects with the FSS step is 160.7 (the range is from 141 (the fifth subject) to 175 (the third subject)) while that of the subjects without the FSS step is 377. It is about 235% increase in the number of adjustable parameters for the classifier with all features. It is believed that using the same parameter training algorithm, the training time is exponentially increased as the number of parameters increases. That is, the time consumed in the parameter training of the classifier with all features is much longer than that of the classifier with the FSS step.

Finally, Table 2 shows the confusion matrix that records the number of recognition errors for all the subjects. Note that, a total of 315 windows for each activity were obtained after the completion of the cross-validation procedure. From the confusion matrix, we can see that the activity “working at a PC” is sometimes confused with “sitting”, because “working at a PC” involves “sitting”. In addition, “brushing teeth” may be misclassified as “running” or “scrubbing” and “vacuuming” may be recognized as ‘walking’, because these activities contain similar movements and frequencies at the upper limbs.

5. Conclusions

A systematic framework for the recognition of human activities using neural networks and acceleration data has been presented in this paper. The acceleration data was collected from a wireless sensing triaxial accelerometer module mounted on the dominant wrist. In general, it is difficult to recognize activities using only one accelerometer. We try to provide a solution to this difficulty via developing an effective design procedure that consists of data pre-processing, feature extraction, efficient feature subset selection, and neural classifier construction. Experiments on successful recognition of eight daily activities with the overall recognition accuracy of 95% have confirmed the effectiveness of the proposed approach. In our future study, we will try to extend the current approach to be capable of determining an optimal number of feature clusters automatically in the feature subset selection, and to collect more acceleration data of a wide range of daily activities for evaluating the new approach. Our ultimate goal is to implement our classifiers in a portable hardware device and to ensure the classifiers can recognize more complex activities accurately in the near future.

References


