Feature Selection and Reduction of Lower Limb Activity Recognition Based on Surface Electromyography and Motion Data

Chunfeng Wei, Hong Wang*, Fo Hu, Jichi Chen, Yanzheng Lu and Yangyang Qi
Department of Mechanical Engineering and Automation, Northeastern University, Shenyang, Liaoning, China
Email: hongwang@mail.neu.edu.cn

Abstract. Daily activity recognition of lower limbs is of great significance to the health care of the elderly and patients with hemiplegia. Surface electromyography (sEMG) signal can directly reflect neuromuscular activity and is an important method for non-invasive monitoring of muscle activity on the body surface. In this paper, a novel method based on sEMG signal and inertial measurement unit (IMU) data to recognize daily activities of lower limbs is proposed. Record sEMG signals and IMU data of fifteen subjects using wearable sensor devices. After pre-processing such as filtering and sliding windows on the data, we extracted seventeen features. A feature selection method based on maximal relevance and minimal redundancy maximal relevance (mRMR) to select representative features. The selected features are input into four machine learning classifiers to classify four daily activities. The performance of the classifier is evaluated using accuracy and receiver operating characteristic curve-area under curve (ROC-AUC) score. The results show that the support vector machine has excellent performance in recognizing the daily activities of human lower limbs.

Keywords. Surface electromyography; feature selection; feature reduction; machine learning; lower limb movement recognition.

1. Introduction
As the population ages and the number of people with lower limb motor dysfunction, it becomes increasingly important to determine lower limb activity in order to maximize quality of life [1]. For this reason, there are three broad approaches to detecting daily activity: computer vision, ambience, and wearable sensors [2]. The computer vision-based method is to detect daily activities by processing videos of human gestures, actions and interactions recorded by cameras. This method tends to interfere with individual privacy and is affected by light and environmental conditions. Ambience sensor-based methods mainly detect daily activities by recognition data recorded by infrared sensors and vibration sensors. This method is easily affected by temperature and other audio signals and noise. Compared with computer vision and ambience methods, wearable sensors have the advantages of not being affected by factors such as light and audio, and high recognition accuracy. Commonly used wearable sensors mainly include sEMG signal sensors and acceleration sensors.

The sEMG signal is the electric potential generated during contraction of skeletal muscle, and it is the electrophysiological response of various activities. Extracting features from sEMG signals has been widely used in daily activities and gesture recognition. For example, Xi et al. [3] used wavelet transform to extract wavelet coherence coefficient features from sEMG signal, and send the selected features to the support vector machine to identify six daily activities. Rabin et al. [4] used short-time Fourier
transform to extract the features of sEMG signals, then the selected features are sent to the K-nearest neighbors classifier to classify six gestures. IMU detect the daily activities and falls by processing acceleration data. Jalloul et al. [5] using IMU to recognition activity at different positions of the individual’s body. Tian et al. [6] using IMU to recognition activity at different positions of the individual’s body. Combining sEMG signal and IMU data to recognize daily activities requires the extraction of a large number of features, which will increase computational cost of the daily activity recognition. Therefore, for daily activity recognition system combining sEMG signal and IMU data, it is necessary to select a small number of features to reduce the amount of calculation while ensuring a high recognition rate.

Based on the above analysis, we proposed a novel daily activity recognition system based on sEMG signal and IMU data feature selection and dimensionality reduction processing. Firstly, pre-process the IMU data and sEMG signals sampled from subjects, such as filtering the data, detecting active segments, and sliding windows. Then, feature extraction is performed on the pre-processed data, and feature selection is performed by the mRMR. Next, the representative features are reduced in dimension through principal component analysis (PCA). Finally, experimental results show that proposed method has great feasibility in daily activity recognition.

2. Materials and Method

2.1. Subjects and Activity
Fifteen healthy subjects were recruited (eight males and seven females; age: 22-28 years old; height: 160-185 cm; weight: 46 -75 kg) to participate in this experiment. They have no history of neurological and muscular diseases. All subjects read and sign an informed consent form approved by the institutional review board. This experiment includes four activities: walking, jumping, upstairs and downstairs. Subjects performed four activities for 30s each in a session, and rested 5s between each activity. Each subject performed 10 sessions, separated by a 1 min rest.

2.2. Data Acquisition and Preprocessing
We use self-made sEMG signals sensor to record seven sEMG signals and three IMU (WT901C485) to record motion data. The sampling frequency of the sEMG signals sensor is 1000 Hz. Seven sEMG signal sensors record the myoelectric signals of the medial femoral muscle, rectus femoris muscle, lateral femoral muscle, semitendinosus muscle, tibialis anterior, lateral gastrocnemius muscle, and medial gastrocnemius muscle. These muscles are closely related to the human limbs activities. The sEMG signal has a bandwidth of 20-150 Hz bandpass filtering and 50 Hz notch filtering. Three IMUs are placed on the chest, outside the thigh, outside the lower leg to record the movement data of the four movements. The sampling frequency of IMU is 100 Hz. A sliding window with a window length of 200ms and an increase of 160ms is used to segment the sEMG signals and IMU data [7].

2.3. Feature Extraction
The feature extraction of sEMG signal mainly includes: time-domain, frequency-domain, time-frequency, and nonlinear dynamic features. The processing method of time-domain features is to treat the sEMG signal as a random signal with a mean value of zero, and the variance varies with the signal strength. The frequency domain feature is that the time domain signal is transformed into a frequency spectrum or a power spectrum. The features extracted in the spectrum or power spectrum are relatively stable, which is conducive to pattern recognition of sEMG signals. Time-frequency characteristics are the characteristics of combining the time domain and frequency domain of the signal. sEMG signal is a kind of nonlinear signal, therefore, nonlinear dynamic characteristics can be extracted for pattern recognition of sEMG signal.

In this work, a total of seventeen features were extracted. As shown in table 1, extracted 14 sEMG signal features [8]. Three features of IMU data extraction are several commonly used mathematical
quantities, namely the average, maximum, and minimum values of triaxial acceleration and the triaxial angular velocity [9].

### Table 1. sEMG features used in this work. Their names, initials, and the parameters used in their extraction are indicated.

| Feature name                  | Abbreviation | Parameters                  |
|-------------------------------|--------------|-----------------------------|
| Autoregressive coefficients   | AR           | 4 order                     |
| Mean absolute value           | MAV          | -                           |
| Root mean square              | RMS          | -                           |
| Integrated sEMG               | IEMG         | -                           |
| Variance                      | VAR          | -                           |
| Simple square integral        | SSI          | -                           |
| Waveform length               | WL           | -                           |
| Willison amplitude            | WAMP         | Threshold=3×10⁻²             |
| Mean frequency                | MNF          | -                           |
| Median frequency              | MDF          | -                           |
| Approximate entropy           | ApEn         | Dimension=2, \( r=0.2\sigma \) |
| Sample entropy                | SpEn         | Dimension=2, \( r=0.2\sigma \) |
| Fuzzy entropy                 | FuEn         | Dimension=2, \( r=0.2\sigma \) |
| Shannon Entropy               | ShEn         | -                           |

### 2.4. Feature Reduction

The mRMR is a filter-based feature selection method. Its core idea is to maximize the correlation between features and categorical variables, and minimize the correlation between features and features. In this paper, for the features extracted above, we use the mRMR feature selection method [10] to select most representative features. In order to balance the accuracy of daily activity recognition and the number of input features, we use PCA to reduce the dimension of features [11].

### 2.5. Classification Method

We send the new feature set obtained by feature selection and dimensionality reduction to the machine learning classifier (linear discriminant analysis (LDA), Boosting Tree, support vector machine (SVM), k nearest neighbor (KNN)) [12]. The kernel function of the LDA classifier is selected as select linear. The boosting method using the decision tree as the basis function is called the boosting tree. The kernel function of SVM classifier chooses the third order polynomial. The KNN classifier Distance metric is the Euclidean distance and equal distance weights are selected. Finally, 10-fold cross-validation is used to evaluate the performance of the four classifiers. The performance of the classifier is evaluated using accuracy and ROC-AUC score. Statistically, a paired samples t-test was performed to compare the performance of classifiers. A significance level of 0.05 was used.

### 3. Result

#### 3.1. Results from Feature Reduction

In this work, we have performed feature selection on two data sets (one with only sEMG signal data and the other is a combined data set of IMU data and sEMG signal), separately and sorted them according to the scores of the features. As shown in figure 1, the accuracy of all classifiers on these two data sets increases as the number of features increases. At the beginning, the accuracy of all classifiers increased rapidly. Analysis showed that features with high scores contributed a lot to the recognition of daily activities. Subsequently, because the low score features contain noise, the accuracy of daily activity recognition increases slowly with the increase in the number of features. When the number of features increases to a certain level, the accuracy of daily activity recognition tends to stabilize. In these two data
sets, the SVM can quickly obtain the highest recognition accuracy, and the accuracy rate of KNN is in the second place, the classification accuracy of Boosting Tree and LDA increases slowly, and the classification accuracy of LDA increases the slowest.

For the sEMG data set, SVM and KNN perform better than LDA and Boosting Tree classification. For the combined data set of IMU data and sEMG signal, the four classifiers achieved high classification performance, and finally, the classification accuracy rate was 99%. The comparison results show that motion data can improve the accuracy of daily activity recognition based on sEMG signals. Synthesize the performance of each classifier on the two data sets, for the sEMG data set, the thirty-four features with the highest score are selected as the input of the daily activity recognition system, and for the data set of the combination of IMU data and sEMG signals, the twenty-five features with the highest score are selected as the input of the daily activity recognition system.

3.2. Classifiers Results

As shown in figure 2, for two data sets, the four classifiers obtain the highest accuracy when using all the extracted features. Compared with the accuracy of all feature input classifiers, the accuracy of the four classifiers is reduced after perform feature extraction, PCA and feature selection combined with PCA processing on the extracted features respectively. Analysis shows that PCA and feature selection inevitably lose some feature information while removing noise, resulting in a decrease in the accuracy of the classifier. The accuracy of LDA and Boosting Tree dropped more and the accuracy of KNN and SVM has not changed significantly.

![Figure 1](image1.png)

**Figure 1.** Accuracy of four classifiers with different number of features, (a) sEMG signal data, (b) IMU data combined with sEMG signal data.

For sEMG signal data, after feature selection and PCA, the accuracy of LDA and Boosting Tree is greatly reduced, which shows that motion data plays a vital role in the recognition of daily activities. After PCA processing the extracted features, SVM has the best classification accuracy (p<0.05). Two types of processing are performed on the extracted features: feature selection and feature selection combined with PCA, respectively. The performance of the classifier is as follows: the classification accuracy of SVM is better than LDA and Boosting Tree (p<0.05), and the classification accuracy of SVM is not significantly different from KNN (p>0.05). For the data set combined with IMU data and sEMG signal, although some feature information will be lost after feature selection and PCA, the accuracy of the classifier is more than 90% and the number of input features is greatly reduced. There are three types of processing for the extracted features: feature selection, PCA and feature selection combined with PCA. The performance of the four classifications is consistent as follows: the classification accuracy of SVM is better than LDA and Boosting Tree (p<0.05). There is no significant difference in classification accuracy between SVM and KNN (p>0.05). For all classifiers, the combined data set of IMU data and sEMG signal has better classification accuracy than sEMG signal data. In the case of the least features, SVM achieves the highest classification accuracy of 97.2%. It means that
method proposed in this paper can yield a better recognize result than the other existential research [13-15].

Figure 2. Accuracy of four classifiers with different feature sets.

As shown in figure 3, the SVM score is the highest, ROC-AUC score of Boosting Tree and LDA are equal, KNN has the lowest score. Therefore, synthesizing the accuracy and ROC-AUC score analysis, based on the mRMR feature selection and PCA technology to process the extracted surface sEMG signal and IMU data features, the SVM performs best in daily activity recognition.

Figure 3. The ROC curves of the four classifiers are based on feature set of feature selection and PCA of extracted sEMG signals and IMU data: (a) LDA, (b) KNN, (c) SVM, (d) Boosting Tree.
4. Conclusion
In this paper, we proposed a novel daily activity recognition method with fewer features to achieve high accuracy. We recorded seven sEMG signals and three IMU data of daily activities from fifteen subjects. After pre-processing, we extracted seventeen features. We use the mRMR feature selection method to select twenty-five representative features and perform PCA on them to further reduce the feature dimension to twelve. The experimental results show that SVM has the best recognition performance, achieving a classification accuracy of 97.2% and ROC-AUV score of 1.

Acknowledgments
We gratefully acknowledge the financial support from the National Key R & D Program of China (2017YFB1300300).

References
[1] Larivière S, Ward N S and Boudrias M-H 2018 Disrupted functional network integrity and flexibility after stroke: Relation to motor impairments Neuroimage-Clinical 19 883-91.
[2] Barshan B and Yurtman A 2020 Classifying daily and sports activities invariantly to the positioning of wearable motion sensor units IEEE Internet Things J. 7 4801-15.
[3] Xi X G, Yang C, Shi J H, Luo Z Z and Zhao Y B 2019 Surface electromyography-based daily activity recognition using wavelet coherence coefficient and support vector machine Neural Process. Lett. 50 2265-80.
[4] Rabin N, Kahlon M, Malayev S and Ratnovsky A 2020 Classification of human hand movements based on EMG signals using nonlinear dimensionality reduction and data fusion techniques Expert Syst. Appl. 149 13.
[5] Jalloul N, Foree F, Viardot G, L’Hostis P and Carrault G 2016 Activity recognition using multiple inertial measurement units IRBM 37 180-6.
[6] Tian Y M, Wang X T, Chen W, Liu Z J and Li L F 2019 Adaptive multiple classifiers fusion for inertial sensor based human activity recognition Cluster Comput. 22 S8141-S54.
[7] Smith L H, Hargrove L J, Lock B A and Kuiken T A 2011 Determining the optimal window length for pattern recognition-based myoelectric control: Balancing the competing effects of classification error and controller delay IEEE Trans. Neural Syst. Rehab. Eng. 19 186-92.
[8] Phinyomark A, Quaine F, Charbonnier S, Serviere C, Tarpin-Bernard F and Laurillau Y 2013 Emg feature evaluation for improving myoelectric pattern recognition robustness Expert Syst. Appl. 40 4832-40.
[9] Zhou B, Wang H, Hu F, Feng N S, Xi H L, Zhang Z H and Tang H 2020 Accurate recognition of lower limb ambulation mode based on surface electromyography and motion data using machine learning Comput. Methods Programs Biomed. 193 11.
[10] Peng H C, Long F H and Ding C 2005 Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy ITPAM 27 1226-38.
[11] Chu J-U, Moon I and Mun M-S 2006 A real-time emg pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand ITBE 53 2232-9.
[12] Subasi A 2013 Classification of EMG signals using pso optimized svm for diagnosis of neuromuscular disorders Comput. Biol. Med. 43 576-86.
[13] Meigal A I, Rissanen S, Tarvainen M P, Karjalainen P A, Iudina-Vassel I A, Airaksinen O and Kankaanpaa M 2009 Novel parameters of surface EMG in patients with parkinson’s disease and healthy young and old controls J. Electromyogr. Kinesiol. 19 E206-E13.
[14] Phinyomark A, Limساกุล C and Phukpattaranont P 2011 Application of wavelet analysis in emg feature extraction for pattern classification Meas. Sci. Rev. 11 45-52.
[15] Ryu J, Lee B H and Kim D H 2017 Semp signal-based lower limb human motion recognition using a top and slope feature extraction algorithm ISPL 24 929-32.